
TESIS DEFENDIDA POR

Víctor Ocyel Chávez Guerrero

Y APROBADA POR EL SIGUIENTE COMITÉ

M.C.E.I.E. María Eugenia Puga Nathal

Director de Tesis

Dr. Humberto Pérez Espinosa

Codirector de Tesis

Dra. María Guadalupe Sánchez Cervantes

Miembro del Comité

Dr. Jesús Ezequiel Molinar Solís

Miembro del Comité

MCC. Ada Mabel Vázquez Paz

*Coordinador del programa de posgrado en
Ciencias de la Computación*

Dra. María Guadalupe Sánchez Cervantes

*Jefe de la División de Estudios
de Posgrado e Investigación*

Agosto de 2021



INSTITUTO TECNOLÓGICO DE CD. GUZMÁN

**TESIS
MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN**

**TEMA DE TESIS:
CLASIFICACIÓN DE COMPORTAMIENTOS DEL
PERRO DOMÉSTICO MEDIANTE VISIÓN
ARTIFICIAL**

**QUE PARA OBTENER EL TÍTULO DE:
MAESTRO EN CIENCIAS DE LA COMPUTACIÓN**

**PRESENTA:
VÍCTOR OCYEL CHÁVEZ GUERRERO**

**DIRECTORES:
M.C.E.I.E. MARÍA EUGENIA PUGA NATHAL
DR. HUMBERTO PÉREZ ESPINOSA**

CD. GUZMÁN JALISCO, MÉXICO, AGOSTO DE 2021



Ciudad Guzmán, **09/agosto/2021**

DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

Asunto: Autorización de impresión de Tesis

**ING. VICTOR OCYEL CHÁVEZ GUERRERO
CANDIDATO AL GRADO DE MAESTRO EN CIENCIAS DE LA COMPUTACIÓN
PRESENTE**

De acuerdo con los Lineamientos para la Operación de los Estudios de Posgrado en el Tecnológico Nacional de México y las disposiciones en este Instituto, habiendo cumplido con todas las indicaciones que la Comisión Revisora realizó con respecto a su Trabajo de Tesis titulado "*Clasificación de comportamientos del perro doméstico mediante visión artificial*", la División de Estudios de Posgrado e Investigación de este Instituto, concede la Autorización para que proceda a la impresión del mismo.

Sin otro particular, quedo de Usted.

ATENTAMENTE

*Excelencia en Educación Tecnológica
"Innova, Transforma y Crea para ser Grande"*

**DRA. MARIA GUADALUPE SÁNCHEZ CERVANTES
JEFA DE LA DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN**



ccp. Archivo
MGSS/megg



Av. Tecnológico No. 100 C.P. 49100 A.P. 150
Cd. Guzmán, Jal. Tel. Conmutador (341) 57520
tecnm.mx jitcg.edu.mx



Resumen de la tesis de **Víctor Ocyel Chávez Guerrero**, presentada como requisito parcial para la obtención del grado de Maestro en Ciencias de la Computación. Zapotlán el Grande, Jalisco, México, Agosto de 2021.

Clasificación de comportamientos del perro doméstico mediante visión artificial..

Resumen aprobado por:

M.C.E.I.E. María Eugenia Puga Nathal

Director de Tesis

Dr. Humberto Pérez Espinosa

Codirector de Tesis

Los perros son los animales de compañía más comunes en todo el mundo, motivados por su excepcional comportamiento social con los humanos. A diferencia de muchos animales, los perros son capaces, entre otras cosas, de aprender órdenes vocales, identificar estados de ánimo, mantener el contacto visual y reconocer expresiones faciales. Además, como tienen una gran agilidad y unos sentidos del olfato y el oído superiores a los de los humanos, los perros han sido entrenados con éxito para tareas cruciales como la búsqueda, el rescate y la asistencia. Por lo tanto, es relevante realizar investigaciones científicas para comprender los fundamentos del comportamiento y los canales de comunicación del perro que aumenten el uso de sus capacidades en beneficio del ser humano, garantizando el bienestar del animal. En este trabajo, fue desarrollado un método computacional para el análisis de las características de comportamiento del perro basado en técnicas de visión artificial. El método propuesto determina el estado emocional del perro en un momento dado, lo que abre un campo prometedor para el desarrollo de nuevas tecnologías que entrenadores y usuarios pueden aprovechar, para mejorar los procesos de selección, entrenamiento y ejecución de las tareas de los perros de servicio. Hemos utilizado una base de datos de videos que incluye la presencia de diferentes estados internos inducidos por varios estímulos. Bajo este método, el mejor valor de precisión que se obtuvo durante las pruebas de clasificación fue de 0.6917 en el mejor modelo entrenado mediante Transfer Learning sobre la arquitectura MobileNet, obteniendo resultados buenos pero no perfectos. El proceso de entrenamiento se realizó con 1067 imágenes distribuidas en cuatro categorías, 343 en la categoría Agresividad, 212 en la categoría Ansiedad, 160 en la categoría Miedo y 352 en la categoría Nada. El método propuesto obtuvo resultados decentes, pero aún puede mejorarse en términos técnicos y metodológicos. Sin embargo, este método puede servir como introducción y sentar las bases para explorar y ampliar el estudio del comportamiento canino utilizando modelos computacionales.

Palabras clave: Visión Artificial, Aprendizaje Máquina, Comportamiento Canino

Abstract of the thesis presented by **Víctor Ocyel Chávez Guerrero**, in partial fulfillment of the requirements for the Master degree in Computer Science. Zapotlán el Grande, Jalisco, México, August 2021.

Classification of domestic dog behaviors using artificial vision.

Abstract approved by:

M.C.E.I.E. María Eugenia Puga Nathal

Director de Tesis

Dr. Humberto Pérez Espinosa

Codirector de Tesis

Dogs are the most common companion animals worldwide, motivated by their exceptional social behavior with humans. Unlike many animals, dogs are able, among other things, to learn vocal commands, identify moods, maintain eye contact, and recognize facial expressions. Besides, since they have great agility and senses of smell and hearing superior to humans, dogs have been successfully trained for crucial tasks such as search, rescue, and assistance. Therefore, it is relevant to carry out scientific research to understand the fundamentals of behavior and communication channels of the dog that increase the use of its capabilities for the benefit of the human being, guaranteeing the welfare of the animal. In this work, we developed a computational method for the analysis of the behavioral characteristics of the dog based on artificial vision techniques. The proposed method determines the dog's emotional state at a given time, which opens a promising field to develop new technologies that trainers and users can take advantage of, to improve the processes of selection, training, and execution of the tasks of service dogs. We used a video database including the presence of different emotional states induced by several stimuli. Using the proposed method, the best test accuracy value we obtained was 0.6917 on the best model trained using Transfer Learning over the architecture MobileNet, getting good but not perfect results. The training process was carried out using 1067 images distributed among four categories, aggressiveness, anxiety, fear and neutral. The proposed method obtained acceptable results but can still be improved in technical and methodology terms. However, this method can be used as an baseline for exploring and expanding the canine behavior study using computational models.

Keywords: Computer Vision, Machine Learning, Canine Behavior

Dedicatoria

A mis mascotas Lana (Q.E.P.D. 2018-2021), Chispita y a todos los perros del mundo

Por ser el motivo de mi inspiración, mis ganas de trabajar en este proyecto y por mostrarme que es posible querer a un nivel tan profundo a otros seres vivos.

La vida cambió muy rápido en muy poco tiempo y aunque sé que Lana o Chispita nunca podrán entender lo que está aquí escrito, espero que aquellos que si puedan comprender estas palabras sepan que a ellas las llevo siempre en el corazón.

Chispita, ojalá vivas para siempre, y sino, que vivas feliz muchos años más.

Lana, espero que hayas sido feliz con nosotros y que sigas siendo feliz allá donde estés.

Agradecimientos

Quiero agradecer a todos los que estuvieron involucrados directa e indirectamente durante el desarrollo de este trabajo, ya que cada uno forma parte importante de mi persona a nivel emocional, profesional e intelectual, y el conjunto de todos estos aspectos dan forma a la persona que soy el día de hoy y alimentan mis ganas de ser una mejor versión de mi el día de mañana para que siempre puedan sentirse orgullosos.

Agradezco a mis papás y mis hermanos por seguir apoyándome a lo largo de los años, tanto en mi formación académica como a nivel personal, han sido un pilar muy importante en mi vida y siempre les estaré agradecido aún cuando me encuentre lejos de ustedes.

A mis amigos más cercanos por siempre estar ahí, tanto en los momentos felices y memorables, así como también en los momentos difíciles, me harían falta un millón de vidas enteras para agradecerles como es debido.

Gracias a mis compañeros y ahora amigos del posgrado por hacer que esta experiencia fuese de lo más divertida y agradable, les deseo lo mejor en sus proyectos de vida.

A mis directores de tesis, el Dr. Humberto Pérez y la maestra María Eugenia por apoyarme con su conocimiento y paciencia a lo largo del desarrollo de este trabajo, gracias por guiarme durante esta etapa tan importante para mi.

Por último pero no por ello menos importante; quiero agradecer a los profesores del ITCG que me dieron clases durante estos años, sus enseñanzas fueron y seguirán siendo importantes en mi desarrollo profesional.

¡Gracias por todo!

Contenido

	Página
Resumen en español	i
Resumen en inglés	ii
Dedicatoria	iii
Agradecimientos	iv
Contenido	v
Lista de figuras	vii
Lista de tablas	ix
I. Introducción	1
I.1. Planteamiento del problema	2
I.2. Objetivos	4
I.2.1. Objetivo general	4
I.2.2. Objetivos específicos	4
I.3. Justificación	4
I.4. Preguntas de investigación	6
I.5. Antecedentes	6
I.6. Aportaciones	10
I.7. Organización de la tesis	11
II. Marco teórico	12
II.1. Aprendizaje máquina	12
II.1.1. Algoritmos de aprendizaje supervisado	13
II.1.2. Redes Neuronales	15
II.1.3. Redes Neuronales Convolucionales	16
II.2. Visión artificial	19
II.2.1. Modelos de colores	20
II.2.2. Segmentación	25
II.2.3. Detección de objetos	28
II.3. Comportamiento animal	30
II.3.1. Comportamiento canino y sus posturas	31
III. Estado del arte	35
III.1. Detección y clasificación de posturas mediante visión artificial . . .	35

Contenido (continuación)

	Página
III.1.1. Visión artificial aplicada a animales	37
III.2. Estudios acerca del comportamiento canino	41
III.2.1. Estudios acerca del comportamiento canino con aplicación de modelos computacionales y visión artificial	42
IV. Metodología	46
IV.1. Corpus de datos	46
IV.2. Pre-procesamiento del corpus	47
IV.3. Generación de muestras	50
IV.3.1. Detección y seguimiento del perro	50
IV.3.2. Aislamiento del perro y recorte automático	52
IV.3.3. Selección de instancias	53
IV.3.4. Algoritmo de sustracción de fondo	55
IV.4. Clasificación manual	56
IV.5. Construcción de vector de características	57
IV.6. Entrenamiento del modelo	58
IV.7. Cálculo de métricas	60
V. Resultados experimentales	62
V.1. Algoritmo SIFT para la comparación y descarte de imágenes	62
V.2. Algoritmos de sustracción de fondo	64
V.3. Arquitecturas y entrenamiento del clasificador	66
V.3.1. Modelo 4 clases	66
V.3.2. Modelo de 3 clases	68
VI. Diseño del prototipo	72
VII. Conclusiones	77
Referencias bibliográficas	81

Lista de figuras

Figura	Página
1. Arquitectura de una red neuronal convolucional simple de 5 capas (O'Shea y Nash (2015))	17
2. Ejemplo de un Kernel 2 x 2	18
3. Modelo de escala de grises	21
4. Modelo RGB	22
5. Modelo CMY	24
6. Modelo HSV representado de forma 2D	25
7. Imagen en escala de grises sin modificar	26
8. Imagen en escala de grises resultante de aplicar un valor de umbral	27
9. Esquema general de actividades	48
10. Detección general de objetos en escena	51
11. Aislamiento en la detección enfocando únicamente al perro	51
12. Recortes almacenados con numeración consecutiva	52
13. Comparación SIFT de imágenes diferentes	63
14. Comparación SIFT de imágenes iguales	63
15. Silueta extraída con el algoritmo de sustracción de fondo	65
16. Problemas para extraer correctamente la silueta del perro después de un movimiento brusco de la cámara	65
17. Sustracción del fondo errónea debido a un cambio repentino en el fondo por un reposicionamiento de la cámara	65
18. Matriz de confusiones modelo de 4 clases	70
19. Matriz de confusiones modelo de 3 clases	71
20. Ventana principal de la aplicación	72
21. Ventana emergente para solicitar acceso a la cámara web	73
22. Cámara web abierta y lista para el análisis	74

Lista de figuras (continuación)

Figura	Página
23. Análisis en tiempo real de una imagen mostrada frente a la cámara . . .	74
24. Análisis de una imagen de la categoría Miedo	75
25. Análisis de una imagen de la categoría Nada	75
26. Análisis de una imagen de la categoría Ansiedad	76

Lista de tablas

Tabla	Página
I. Número de videos en cada categoría de acuerdo a la revisión y clasificación manual.	50
II. Lista de conductas a observar	57
III. Resultados por modelo del entrenamiento de clasificadores de 4 clases	68
IV. Comparación de precisión por clase en los modelos de 4 clases y de 3 clases	69
V. Porcentaje de aciertos durante la prueba de clasificación: modelo 4 clases	71
VI. Porcentaje de aciertos durante la prueba de clasificación: modelo 3 clases	71

Capítulo I

Introducción

El estudio del comportamiento del perro es una tarea que, al día de hoy, se ha convertido en parte importante en la sociedad, ya que algunos perros tienden a ser candidatos para ser perros de asistencia o servicio, ya sean perros de rescate, perros guías, perros guardianes, entre otros. La evaluación y el análisis del comportamiento se lleva a cabo de diversas maneras, ya sea con cuestionarios a los dueños de los perros, evaluación a partir de expertos, medidas estandarizadas y estudios de observación. Sin embargo, debido a que en ocasiones es posible que algún detalle se escape o que alguna variación no sea detectada por el humano, es necesario generar o crear alguna herramienta que sirva como auxiliar en la evaluación exhaustiva del comportamiento de los perros ya que una computadora puede analizar todo un día de actividades del perro y para un humano es complicado. Esto con el fin de capturar esos detalles y reportarlos a los expertos, los cuales tomarán una decisión más completa ya que contarán con más información respecto al comportamiento de los perros. En el presente proyecto se pretende desarrollar una herramienta de análisis de video para evaluar características y comportamiento del perro, más específicamente, perros que son candidatos o perros que ya están en proceso de entrenamiento para realizar labores de rescate, teniendo así un impacto en la sociedad, auxiliando a los entrenadores de perros y aportando una herramienta al campo de las ciencias computacionales en materia de análisis de comportamiento animal. A continuación, se mencionan antecedentes relacionados con el protocolo de tesis, la justificación, preguntas de investigación tanto principales como

auxiliares, así como la hipótesis planteada para este proyecto.

I.1. Planteamiento del problema

Los perros son los animales de compañía más comunes en todo el mundo, llegando a ocupar un lugar relevante para la sociedad humana. Por lo tanto, resulta interesante el estudio etológico de esta especie. Una mejor comprensión del comportamiento y expresiones del perro permitirá tomar decisiones más acertadas para mejorar el bienestar del perro y para maximizar el aprovechamiento de sus comportamientos y habilidades en beneficio del ser humano. Los perros tienen varios comportamientos característicos con los cuales transmiten información hacia sus conespecíficos y heteroespecíficos, por ejemplo, mover la cola, mostrar los colmillos, ladrar, adoptar alguna postura corporal (Handelman (2012)). El principal factor regulador detrás de estos comportamientos del perro es el estado interno del animal (Dugatkin (2020)), al investigar estas señales externas, se podrá comprender mejor el origen, motivo y función de las formas de expresión estudiadas. El presente proyecto tiene como objetivo la investigación de las diferencias expresivas entre diferentes estados internos del perro. Es de interés comprender mejor el lenguaje corporal y vocal usado por los perros para comunicarse y desarrollar tecnología que nos asista en este sentido. Así mismo, se desea saber qué expresiones particulares están relacionadas con los estados emocionales del perro. Se propone el análisis automático de videos de expresiones del perro como una nueva herramienta en la evaluación del estado interno del perro. Tales estados que generan expresiones corporales, y faciales ofrecen un campo prometedor para desarrollar nuevas tecnologías para la interpretación automática de las expresiones caninas. Para el desarrollo del proyecto, se utilizará una base de datos de información representativa

de las expresiones de los perros en diferentes estados internos. La información se analizará en video, además de información contextual y de caracterización de los perros registrados. Se hará una investigación de los factores principales que afectan los estados internos de los perros. Posteriormente se desarrollará un método para el análisis de expresiones caninas basado en técnicas de inteligencia artificial de última generación para generar una interpretación.

Actualmente analizar el comportamiento de los perros mediante las técnicas convencionales implica que la persona evaluadora tenga un conocimiento relativamente amplio del comportamiento animal. Además de que muchas veces pueden pasar por alto algunos detalles importantes o la evaluación conlleva implicaciones personales del observador. Es decir, cuestiones que el observador evalúa sin un criterio estandarizado y con adjetivos de evaluación tales como “curioso” (Uher y Asendorpf (2008)), además de que una persona no puede estar 24 horas observando al perro y en algún lapso de tiempo de esas 24 horas, el animal puede presentar algún cambio de comportamiento o realizar alguna acción que sea relevante para la evaluación pero que pasa desapercibida debido a la ausencia del evaluador. El proyecto se centrará en perros candidatos a realizar tareas de apoyo, más concretamente a perros de rescate, debido a que gran parte de México se encuentra en una zona altamente sísmica (Pozos-Estrada *et al.* (2014)), constantemente se presentan sismos y temblores a lo largo del país, algunos resultando en derrumbes de edificios o complejos habitacionales donde las labores de rescate deben iniciar inmediatamente para minimizar la pérdida de vidas humanas, siendo una carrera contra el tiempo donde los perros de rescate juegan un papel muy importante para encontrar a personas que se encuentran atrapadas en los escombros.

I.2. Objetivos

A continuación se presentan los objetivos de este trabajo de tesis.

I.2.1. Objetivo general

Desarrollar un sistema de monitorización y evaluación del comportamiento del perro mediante el análisis automático de videos que provea información útil a los entrenadores de perros de rescate.

I.2.2. Objetivos específicos

Los objetivos específicos que coadyuvan al cumplimiento del objetivo general son los siguientes:

- Investigar las diferencias expresivas entre diferentes estados internos del perro.
- Definir los movimientos y posturas a clasificar.
- Desarrollar clasificadores de posturas y/o movimientos del perro.
- Diseñar y desarrollar un sistema de soporte a los entrenadores de perros de rescate.
- Evaluar el desempeño del sistema.

I.3. Justificación

Los perros son de los animales de compañía más comunes en el mundo, por lo tanto, son de mucha importancia para gran parte de la sociedad. Con el desarrollo del presente proyecto se pretende crear un sistema de análisis automático de video de expresiones del

perro domestico para auxiliar en la comprensión de los estados de ánimo presentes en un perro, ya que es un ser vivo y su estado emocional se ve afectado por factores tanto del entorno que lo rodea como situaciones propias del animal. Tomando en cuenta la importancia de estos animales en la sociedad se vuelve un factor de impacto conocer en todo momento el comportamiento que puede tener un perro debido a que además de ser simples mascotas llevan a cabo labores auxiliares del día a día como es rescates, perros guías, perros de seguridad, y obviamente no todo perro es apto para desempeñarse en estos rubros debido a las características y personalidad de cada individuo.

Se pretende que el desarrollo de este proyecto sirva como una nueva tecnología en apoyo para la comprensión del estado de salud o emocional de los perros para poder tomar decisiones más oportunas acerca de cuáles perros son los más calificados para llevar a cabo labores de apoyo para nuestra sociedad, ya que además de ser animales auxiliares en ciertos ámbitos, son seres vivos y compañeros con sus propios estados de ánimo, carácter y rasgos físicos que deben ser tomados en cuenta. Además, desde el punto de vista computacional desarrollar sistemas de aprendizaje máquina y reconocimiento de características en animales abre las posibilidades para que esta tecnología sea utilizada también en personas en un futuro no muy lejano. Desarrollar estos sistemas permite aportar al área de la visión artificial nuevas herramientas para el desarrollo de nuevas tecnologías computacionales y apoyar en futuras investigaciones que tengan como base el análisis de imágenes o video, bien sea enfocado en perros como en algún otro animal, incluso llegando a utilizarse con personas.

I.4. Preguntas de investigación

Con base a los objetivos planteados, se formularon la siguientes preguntas de investigación:

- ¿Cómo el sistema de detección impacta en la evaluación, caracterización y rasgos del comportamiento de un perro de rescate?
- ¿Qué tipo de características visuales son las más adecuadas para capturar los movimientos del perro?
- ¿Qué algoritmos de aprendizaje máquina generan los mejores resultados en este tipo de tareas?

I.5. Antecedentes

Diversas investigaciones arrojan resultados sobre el monitoreo del comportamiento de animales, Tillett et al. (Tillett *et al.* (1997)) desarrollan sistemas capaces de seguir a animales de granja a través de sensores que están directamente unidos a los animales, lo cual muchas veces afecta el comportamiento del animal ya que perciben el dispositivo; además de que estos sistemas únicamente registran la posición de los dispositivos.

Ante esta problemática los autores sugirieron el uso de cámaras para captar información visual y a partir de los datos obtenidos desarrollar un modelo computacional basado en imágenes para seguir el movimiento de los animales. Es aquí donde se da entrada al registro del comportamiento de los animales a través de cámaras y sistemas computacionales.

Más adelante, en años más recientes algunos autores han trabajado para el reconocimiento de posturas caninas, como es el caso del trabajo realizado por Ribeiro

(Ribeiro *et al.* (2009)) este trabajo está enfocado en el reconocimiento de posturas como estar de pie, sentado, caminando o acostado mediante el uso de sensores acelerómetros ubicados en la parte dorsal y lumbar del perro mediante el uso de un arnés especial, generando un modelo computacional con la información obtenida de los acelerómetros a partir de un esquema base predeterminado de movimientos propios de cada postura, este trabajo estuvo orientado a perros que llevan tareas de encontrar personas en labores de rescate durante emergencias y se probó con los perros de la policía provincial de Ontario, Canadá.

Y por otro lado, llevando un enfoque para el entrenamiento de perros, los autores Brugarolas *et al.* (Brugarolas *et al.* (2013)) desarrollaron algoritmos basados en aprendizaje máquina enfocados a identificar posturas específicas de perros tales como estar sentado, de pie, acostado, de pie en dos patas y comiendo del suelo a partir de un análisis del cambio en el movimiento de los perros, es decir, un análisis inercial, esto se logró combinando un sistema de sensores inalámbricos, de entre los cuales el artículo destaca un acelerómetro CMA3000 y un giroscopio L3G4200D conectados a un chip CC2540 de Texas Instruments que hace la función de transmisor bluetooth, los datos recolectados de estos sensores se envían a una computadora donde se ejecuta MATLAB. Los autores desarrollaron un modelo de movimiento basado en estos datos inerciales, y posteriormente para clasificar las posturas utilizaron algoritmos de Machine Learning implementados en la herramienta WEKA, enfocándose principalmente en los algoritmos de Random Forest, K vecinos más próximos y árbol de modelo logístico.

Las posturas de los perros no ha sido el único campo que se ha estudiado, sino también su comportamiento durante alguna emergencia, ya que en alguna situación inesperada, los perros tienden a presentar acciones que, bajo condiciones normales, no mostraría, por ejemplo, puede que empiecen a correr, derriben a las personas que

se encuentren cerca o por el contrario no actuar ante la emergencia (Robinson *et al.* (2015)) por lo que los autores Robinson *et al.* mencionan que resulta de interés realizar un monitoreo de los patrones conductuales de los perros durante una situación de emergencia ya que los perros pueden reaccionar no solo mediante posturas o movimiento, sino también mediante vocalizaciones irregulares que indiquen estrés o una afección negativa y usando estos datos se puede desarrollar diversas tecnologías con un amplio abanico de aplicaciones.

En otros trabajos como el de Mealin *et al.* (Mealin *et al.* (2016a)) se propone utilizar tecnología que nos permite obtener mayor información visual utilizando el dispositivo de Kinect, el cual posee una cámara infrarroja, una cámara de profundidad y una cámara regular; haciendo uso de la cámara infrarroja y la cámara regular se capturan varias secuencias de videos específicamente de perros para analizar y clasificar mediante un modelo semi-supervisado las posturas de los perros, siendo que se evaluaban las posturas de estar de pie o pararse, sentarse y acostarse, lo cual es interesante ya que está relacionado al trabajo que se pretende desarrollar.

Otro trabajo relacionado de los autores Mealin *et al.* (Mealin *et al.* (2016b)) analizaron imágenes con la finalidad de aislar y detectar caninos por sobre otros objetos en el escenario mediante un algoritmo no supervisado, utilizando una vez más el Kinect y haciendo uso de la cámara de profundidad y la cámara infrarroja. La combinación de ambos trabajos podría resultar en un sistema capaz de; primero capturar video de un perro, después preparar un segmento de video aislando al perro dentro del video para posteriormente analizar la secuencia de imágenes para detectar las posturas básicas ya mencionadas. El uso de tecnologías como el Kinect ofrece más oportunidades al momento de trabajar con visión computacional y procesamiento de imágenes digitales, tanto al trabajar con imágenes estáticas o con videos, sin embargo, se requiere una

capacidad de cómputo mayor para procesar las imágenes de profundidad e infrarrojas, además de que los procedimientos para analizar dichas imágenes es más complejo que utilizando imágenes RGB comunes, eso sin mencionar que es necesario adquirir el kit de desarrollo de Kinect. Dicho esto, a lo largo de este trabajo se muestra que es posible realizar la detección y clasificación de comportamientos de perros utilizando cámaras convencionales e imágenes RGB, lo cuál es una ventaja ya que no requiere un poder de cómputo tan elevado ni dispositivos adicionales.

Por otra parte, en el trabajo presentado por Amir et al. (Amir *et al.* (2017)) se propone otra metodología para analizar comportamiento de perros domésticos. Se presenta una herramienta de nombre K-9 Blyzer Tool, desarrollada en OpenCV, una librería para el lenguaje de programación Python; con esta herramienta se analiza el comportamiento que tiene un perro al interactuar con un objeto, en este caso un robot pequeño que se mueve haciendo ciertos sonidos para generar cierta interacción con el perro y evaluando comportamientos del animal como es tocar o interactuar con el robot, acercarse al robot y alejarse de él, esto se calcula a partir de la distancia relativa entre el perro y el objeto de interacción mediante un umbral predefinido por los autores y posteriormente se etiquetan los frames que correspondan a dichos cálculos. Lo interesante es que no se requiere de una cámara especializada como el Kinect para la captura del video y el lenguaje de desarrollo es sencillo de aprender e implementar.

Ehsani et al. (Ehsani *et al.* (2018)) presentan un trabajo donde a partir de sensores y cámaras acopladas a un perro se modela el comportamiento de un perro, extrapolando los movimientos registrados a un modelo 3D utilizando técnicas de digitalización para estudiar el posible comportamiento y las acciones que podría llevar a cabo un perro a la hora de realizar ciertas actividades; este análisis requiere imágenes previas capturadas en video mediante cámaras montadas con arneses al perro, además de recopilar información

sobre la posición del cuerpo del perro y el movimiento que genera durante la grabación de dicho material, posteriormente a esto se analizan los videos y los datos mediante un modelo computacional, una vez finalizado el análisis, el modelo recrea los movimientos probables que llevará a cabo al realizar ciertas acciones o reaccionar ante algunos estímulos.

En resumen, existen múltiples herramientas de captura de video, técnicas y tecnologías de desarrollo para generar sistemas capaces de analizar patrones de comportamiento o de postura en video, y aprovechando todas estas opciones se puede diseñar un algoritmo más robusto combinando las mejores características de los trabajos ya mencionados. La diferencia del presente proyecto es que se desea hacer no solo un seguimiento visual de un perro mediante una cámara, sino que también se pretende hacer un análisis conductual en base a posturas, comportamientos y acciones que lleve a cabo el canino a lo largo del día, identificando además de las posturas básicas como estar de pie, sentado o acostado como en los trabajos ya mencionados, se quiere analizar detalles más específicos que denoten información relevante para los expertos en el entrenamiento de perros de servicio y de rescate.

I.6. Aportaciones

Las principales contribuciones de esta tesis son las siguientes:

- Un sistema de análisis y clasificación de comportamientos enfocado en perros que permita obtener información detallada respecto al comportamiento general de un perro, esto con el fin de auxiliar a los entrenadores de perros de rescate a tomar decisiones respecto a que perros pueden ser o no de utilidad teniendo más datos disponibles.

I.7. Organización de la tesis

El resto de la tesis se organizará de la siguiente manera:

- Marco teórico
- Estado del arte
- Metodología
- Resultados experimentales
- Diseño del prototipo
- Conclusiones y trabajo futuro
- Referencias bibliográficas

Capítulo II

Marco teórico

En este capítulo se muestra la base teórica en la que se fundamenta el presente proyecto de tesis.

II.1. Aprendizaje máquina

El aprendizaje máquina es la disciplina que básicamente se enfoca en dos interrogantes que están totalmente relacionadas entre sí (Jordan y Mitchell (2015)):

- ¿Cómo es posible construir sistemas computacionales que automáticamente mejoren a través de la experiencia de ejecución?
- ¿Cuáles son las leyes fundamentales que gobiernan todos los sistemas de aprendizaje incluyendo computadoras, humanos y organizaciones?

Los estudios acerca del aprendizaje máquina han sido fundamentales hoy en día ya que pasaron de ser simples estudios por curiosidad en un laboratorio a ser tecnologías prácticas aplicadas ampliamente alrededor del mundo. Tomemos como ejemplo el desarrollo de inteligencia artificial. El aprendizaje máquina se ha elegido como el método predilecto para el desarrollo de software como visión computacional, reconocimiento de voz, procesamiento de lenguaje natural, control de robots, entre otras aplicaciones. Sin embargo, la conceptualización de aprendizaje máquina no es algo reciente, por ejemplo, Donald Michie menciona que sería muy útil si las computadoras pudieran aprender de la experiencia y con ello, automáticamente mejorar la eficiencia de sus propios programas

durante la ejecución (Michie (1968)). La teoría del aprendizaje máquina viene dada del interés de entender y aplicar la forma en que las formas de vida aprenden, entendiendo como los sistemas cognitivos de los seres vivos han pasado por adaptación y evolución a través de billones de años. Los modelos naturales de aprendizaje y adaptación son representaciones de procedimientos y estructuras neuronales robustas. El desarrollo del aprendizaje máquina tiene especial interés en las redes neuronales (Goldberg y Holland (1988)), representaciones artificiales del cerebro, principalmente tratando de imitar el funcionamiento del cerebro humano. Un problema de aprendizaje puede ser definido como el problema de mejorar alguna medida de rendimiento cuando se ejecuta alguna tarea, como si se tratara de algún tipo de experiencia por entrenamiento. Por ejemplo, el aprendizaje para la tarea de detectar si una transacción bancaria con una tarjeta de crédito es o no un fraude la medida de rendimiento a mejorar sería la precisión del clasificador para detectar esa operación, esto lo haría en base a la experiencia de situaciones previas con otras transacciones de tarjetas de crédito que hayan sido marcadas como fraudulentas o no fraudulentas (Jordan y Mitchell (2015)).

II.1.1. Algoritmos de aprendizaje supervisado

Normalmente los algoritmos de aprendizaje máquina están organizados en ciertas categorías basándose en los resultados finales que desean obtenerse, y una de esas categorías son los algoritmos de aprendizaje supervisado, los cuales tienen como finalidad generar una función que sea capaz de mapear entradas y convertirlas en las salidas deseadas. El aprendizaje máquina supervisado está descrito como la búsqueda de algoritmos que sean capaces de realizar predicciones a partir de hipótesis generales previamente aprendidas, esto tiene un gran campo de aplicación en el desarrollo de

sistemas inteligentes y de clasificadores (Osisanwo *et al.* (2017)).

De acuerdo a (Ayodele (2010)) los algoritmos de aprendizaje supervisado pueden clasificarse en las siguientes categorías:

- Clasificadores lineales
- Regresión Logística
- Clasificador Naïve Bayes
- Perceptrón
- Máquinas de vectores de soporte (SVM por sus siglas en inglés)
- Clasificadores Cuadráticos
- K-Means Clustering
- Boosting
- Árboles de decisión
- Bosques aleatorios (RF por sus siglas en inglés)
- Redes neuronales
- Redes Bayesianas
- Entre otros.

Si bien la variedad de algoritmos es bastante amplia y cada cual con sus propias características, en este trabajo nos enfocaremos en las Redes Neuronales, ya que pueden realizar una serie de tareas de regresión y / o clasificación a la vez, aunque comúnmente

cada red realiza solo una de ambas (Bishop *et al.* (1995)) además de que las redes neuronales han mostrado un excelente desempeño en el análisis de imágenes y videos (O'Shea y Nash (2015)).

II.1.2. Redes Neuronales

Las Redes Neuronales Artificiales (ANN por sus siglas en inglés) son sistemas computacionales fuertemente inspirados en la forma en que funcionan los sistemas nerviosos biológicos, como por ejemplo el cerebro humano. La premisa principal de las redes neuronales artificiales es un único sistema interconectado a múltiples nodos interconectados, los cuales reciben el nombre particular de neuronas, ya que su tarea fundamental es aprender de manera conjunta de las entradas que recibe para obtener una salida final (O'Shea y Nash (2015)).

Las redes neuronales o redes neuronales artificiales dependen de ciertos aspectos fundamentales (Osisanwo *et al.* (2017)), los cuales son las funciones de entrada y de activación, la arquitectura de la red y los pesos de cada conexión de entrada.

Una vez que se establecen los dos primeros aspectos, es decir, las funciones de entrada y activación, el posterior comportamiento de la red se definirá por los valores actuales de los pesos de cada entrada. Estos pesos suelen ser aleatorios inicialmente durante el entrenamiento, posteriormente y conforme avanza el entrenamiento de la red se compara la salida obtenida en esa iteración del entrenamiento contra la salida real o deseada, así los pesos se van ajustando ligeramente para tratar de obtener los valores de salida más cercanos a la salida deseada.

A su vez, de las redes neuronales artificiales se desprenden las Redes Neuronales Convolucionales (CNNs por sus siglas en inglés), las cuales en teoría son análogas a

las ANN en el sentido de que también están formadas por neuronas interconectadas con el propósito de auto optimizarse a través del aprendizaje, cada neurona recibe una entrada específica y realiza una operación con esa entrada, la diferencia principal entre una y otra es que las redes convolucionales son usadas principalmente en el campo de reconocimiento de patrones con imágenes, por lo que si se trabaja con características procedentes de imágenes permite reducir el número de parámetros requeridos para configurar el modelo.

II.1.3. Redes Neuronales Convolucionales

El principal enfoque de las redes neuronales convolucionales es que las entradas a ser procesadas estén conformadas principalmente de imágenes (O'Shea y Nash (2015)) lo cual permite enfocar la arquitectura de la red neuronal de forma que resulte mejor lidiar y trabajar con este tipo específico de datos.

La arquitectura general de una red neuronal convolucional está conformada por tres tipos de capas (layers), capas convolucionales (convolutional layers), pooling layers y capas completamente conectadas (fully connected layers). Una vez que las tres capas estén reunidas se puede considerar que se ha formado la arquitectura base de una red neuronal convolucional.

Cada capa mencionada previamente tiene un trabajo o una función específica, las cuales serán descritas brevemente a continuación:

- Las capas de entrada mantendrán los valores de los píxeles de cada imagen
- Las capas convolucionales determinará la salida de las neuronas, las cuales están conectadas a regiones locales de la entrada, todo esto mediante el cálculo del producto escalar entre sus pesos y la región conectada al volumen de entrada.

Además, la unidad lineal rectificada (comúnmente abreviada a ReLu) tiene como objetivo aplicar una función de activación "por elementos", como sigmoide, a la salida de la activación producida por la capa anterior.

- Las Pooling Layers simplemente realizará una reducción de resolución a lo largo de la dimensionalidad espacial de la entrada dada, reduciendo aún más el número de parámetros dentro de esa activación.
- Las capas completamente conectadas serán las encargadas de producir puntuaciones o scores de cada clase que se utilizarán para la clasificación.

Así pues, las redes neuronales convolucionales son muy útiles a la hora de tratar principalmente con imágenes, pero funcionando esencialmente como una red neuronal artificial convencional con algunas pequeñas diferencias.

En la figura 1 se puede observar un ejemplo sencillo de una red neuronal convolucional simple.

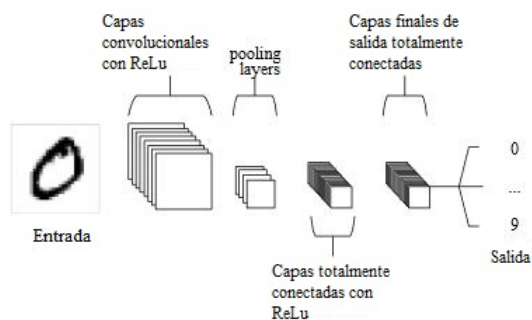


Figura 1. Arquitectura de una red neuronal convolucional simple de 5 capas (O'Shea y Nash (2015))

En otras palabras una red convolucional consiste en un set finito de capas de procesamiento que son capaces de aprender varias características a múltiples niveles de abstracción a partir de datos de entrada, como por ejemplo una imagen, la cual se

descompone conforme va atravesando las capas internas de la red convolucional (Ghosh *et al.* (2020)). Las capas iniciales de la red son capaces de extraer características de alto nivel o de baja abstracción, mientras que las capas más profundas son las que pueden extraer características de bajo nivel o con un alto nivel de abstracción, terminando la extracción de características en una capa final totalmente conectada, esta capa es la encargada de, a partir de la información y las características extraídas en las capas internas previas, analizar dicha información, formar un mapa de características y así llevar a cabo la tarea final para la que esté diseñada dicha red.

Las redes convolucionales hacen uso de algo llamado Kernel, el cual básicamente se puede describir como una matriz de valores discretos o números, donde cada valor dentro de dichas matrices pasan a ser conocidos como los pesos de dicho Kernel. Al iniciar el proceso de entrenamiento de un modelo convolucional, todos o la mayoría de los pesos del kernel se inician con valores aleatorios, aunque este no siempre es el caso, siendo posible asignar los pesos de manera manual bajo ciertas condiciones o de inicializar dichos pesos bajo cierto patrón preestablecido (Ghosh *et al.* (2020)). Una vez que el proceso de entrenamiento comienza a avanzar, después de cada época los pesos se ajustan y el kernel aprende a extraer características con cierto nivel de abstracción. En la figura 2 se puede apreciar un ejemplo sencillo de un Kernel con un tamaño de 2 x 2.

0	1
-1	2

Figura 2. Ejemplo de un Kernel 2 x 2

Actualmente las redes neuronales convolucionales son especialmente útiles además

de ser capaces de alcanzar resultados prometedores en varias aplicaciones en el campo de la visión computacional como la clasificación de imágenes, detección de objetos, detección de rostros, reconocimiento de vehículos, de expresiones faciales, de texto y muchos más, pero las redes neuronales convolucionales no solo están limitadas a aplicaciones de visión artificial, siendo también especialmente útiles en aplicaciones como reconocimiento de voz, análisis y clasificación de audio, entre otros (Ghosh *et al.* (2020)).

II.2. Visión artificial

Primeramente la visión es definida por D. Man como: “un proceso que produce a partir de las imágenes del mundo exterior una descripción que es útil para el observador y que no tiene información relevante” (Man y Vision (1982)).

Ahora, tomando en cuenta esta definición, podemos definir la visión computacional como un proceso en el que se desea extraer características de una imagen o una serie de imágenes para su descripción e interpretación por una computadora, despreciando la información que no es útil y eliminando fallas en la imagen para un mejor análisis (Sucar y Gómez (2011)).

La visión computacional actualmente tiene muchas aplicaciones en muchos campos de la ciencia y la tecnología, mencionando solo como algunos ejemplos muy generales:

- Robótica,
- Manufactura,
- Medicina,
- Interpretación de escritura, dibujos, planos, etc.,

- Análisis de imágenes para astronomía,
- Interpretación de imágenes aéreas y de satélites, entre otros.

Una imagen se forma cuando un “sensor”, sea el ojo de una persona o una cámara de una computadora detecta y procesa la interacción y refracción de la luz con los objetos del entorno. A grandes rasgos, una imagen es el análisis y la interpretación de la información recibida a través de un sensor, ya sea biológico o digital. De ahora en adelante solo serán considerados los sistemas digitales y lo concerniente a la visión computacional y la interpretación de las imágenes digitales. Una imagen digital es una imagen en la cual los valores de intensidad de la refracción de la luz y su ubicación espacial ha sido discretizada mediante una función con un arreglo bidimensional de funciones delta, donde cada valor de intensidad es llevada a una matriz de números discretos, por ejemplo los valores de los colores RGB en las computadoras van desde 0 hasta 255 (Sucar y Gómez (2011)).

II.2.1. Modelos de colores

Al momento de trabajar con imágenes en el ámbito de la visión artificial nos encontramos con diversos modelos de colores, cada uno con sus respectivas ventajas y desventajas, donde cada modelo cuenta con sus propias características acerca del manejo de cada color en el espectro visible. Es importante conocer al menos los conceptos básicos del manejo de colores a la hora de trabajar con un sistema de visión artificial, ya que, precisamente, una imagen tendrá que ser manipulada en alguno de estos modelos, ya sea para hacer una corrección de color, algún cambio de color, cambio de modelo, entre otros. Entre los principales modelos aplicables en el ámbito de la computación podemos encontrar (Ibraheem *et al.* (2012)):

- Escala de grises
- RGB (Rojo, Verde, Azul)
- CMY (Cian, Magenta, Amarillo) con variaciones de CMYK o CMYB
- HSV también llamado HSB (Tonalidad o matiz, Saturación, Brillo o Valor)

A continuación se hará una breve descripción de cada modelo de color.

Modelo de escala de grises

Este es un modelo de colores relativamente sencillo, donde el color de cada píxel en una imagen está representado por un valor entre 0 y 255, donde el valor 0 significa ausencia total de color lo que resulta en color negro y 255 la tonalidad máxima que es el color blanco, así pues, todos los valores comprendidos entre 1 y 254 son tonalidades de grises desde los más oscuros hasta los más claros como se puede observar en la figura

3.



Figura 3. Modelo de escala de grises

Modelo de color RGB

El modelo de color RGB (Red, Green, Blue por sus siglas en inglés) utiliza la mezcla de los colores rojo, verde y azul para crear los colores restantes del espectro visible por

el ojo humano, y particularmente en el área de la computación los valores de cada color primario van desde el 0 hasta el 255, siendo el 0 la ausencia total del color en cuestión hasta el 255 que representa el valor máximo de la tonalidad de ese color, por ejemplo, la combinación (0,0,0) representa que existe un valor de 0 en el canal rojo, un valor de 0 en el canal verde y un valor de 0 en el canal azul, por lo que el color resultante es el negro o ausencia total de color, mientras que una combinación (255,255,255) representa la máxima tonalidad del color rojo, la máxima tonalidad del color verde y la máxima tonalidad del color resultando en el color blanco, así pues las diversas combinaciones entre los colores primarios resultarán en los demás colores del espectro visible como puede ser el amarillo, el morado, el naranja, entre muchos otros. En la figura 4 se puede observar un ejemplo del modelo RGB.

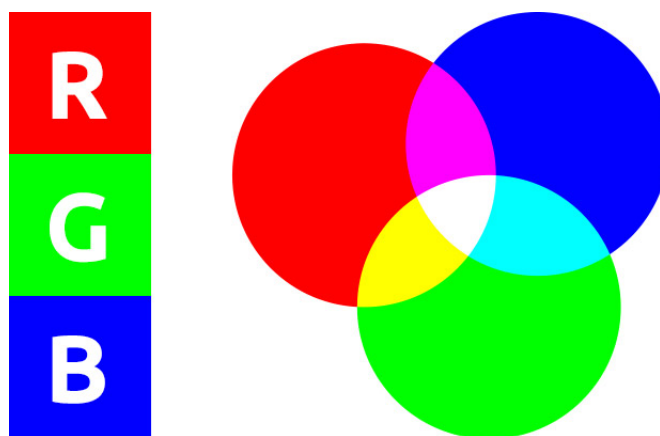


Figura 4. Modelo RGB

Modelo CMY

El modelo CMY (Cyan, Magenta, Yellow) es un modelo substractivo en base al sistema RGB, lo que significa que en este modelo la presencia de todos los colores da como resultado el negro y la ausencia de estos el color blanco, esto se puede observar en

la figura 5, sin embargo existe un modelo alternativo en donde el color negro se agrega de forma independiente, formando parte de la combinación final y tomando el nombre de CMYK o CMYB (Cyan, Magenta, Yellow, Key (Black)). Los valores de estos colores se forman a partir de los valores provenientes del modelo RGB mediante operaciones sencillas, siendo las siguientes.

- $\text{Cian} = 1 - \text{Rojo}$
- $\text{Magenta} = 1 - \text{Verde}$
- $\text{Amarillo} = 1 - \text{Azul}$

Y para convertir del modelo CMY a CMYK se realizan las siguientes operaciones (Ibraheem *et al.* (2012)):

- $\text{Negro} = \text{mínimo}(\text{C}, \text{M}, \text{A})$
- $\text{Cian} = (\text{Cian} - \text{Negro}) / (1 - \text{Negro})$
- $\text{Magenta} = (\text{Magenta} - \text{Negro}) / (1 - \text{Negro})$
- $\text{Amarillo} = (\text{Amarillo} - \text{Negro}) / (1 - \text{Negro})$

Modelo de color HSV o HSB

El modelo HSV (Hue, Saturation, Value), HSB (Hue, Saturation, Brightness) es un modelo que realiza una transformación no lineal del modelo RGB. Este modelo es usado principalmente en el análisis de imágenes en el proceso de segmentación (Ibraheem *et al.* (2012)). Para obtener un color en este modelo es necesario combinar los tres de sus componentes:

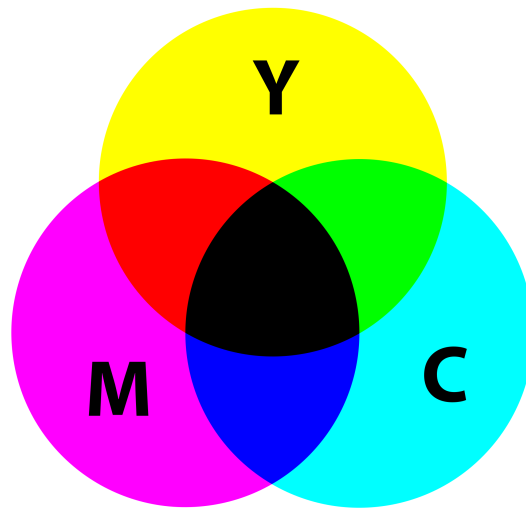


Figura 5. Modelo CMY

- H: Hue (matiz) se maneja como un círculo o como un conjunto de grados que van desde 0° hasta 360° , donde $0^\circ = \text{rojo}$ al igual que 360° , $120^\circ = \text{verde}$ y $240^\circ = \text{azul}$, y para los colores mixtos se utilizan los valores o grados intermedios entre cada uno de los valores mencionados.
- S: El valor de la saturación se maneja en valores de porcentaje, es decir, de 0 a 100 por ciento, este valor representa que tanta tonalidad tendrá de cierto color, por ejemplo, entre más cercano se encuentre el valor al 0, el color tendrá una tonalidad grisácea y decolorada, mientras que al contrario, entre más cercano al 100, el color será más claro o más púro.
- V/B: Value/Brighthness (valor/brillo) también se maneja en valores de porcentaje, es decir, de 0 a 100 por ciento. Este parámetro corresponde al eje de blanco-negro, por lo tanto, un valor de 0 representa un color negro total, mientras que un valor 100 representa un color totalmente blanco o con más "brillo"(saturado en color blanco).

Se puede observar la representación del modelo HSV en la figura 6 en su forma 2D.

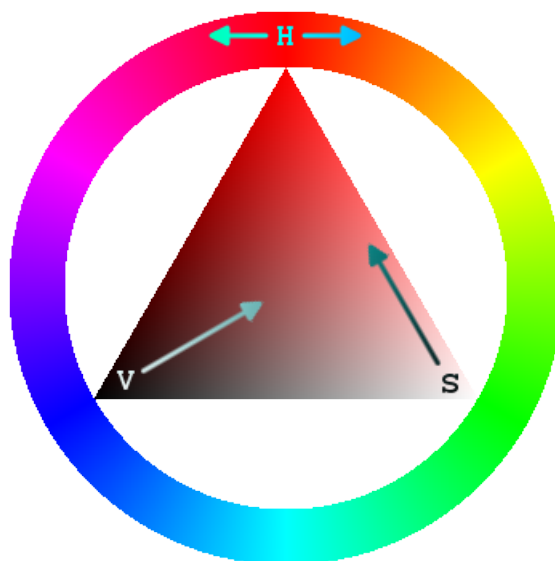


Figura 6. Modelo HSV representado de forma 2D

Así pues, entender los fundamentos básicos de cada modelo de colores es necesario para diversos procesos en el campo de la visión artificial como el mapeo de colores de una imagen particular, la implementación de métodos del valor umbral, métodos de detección de bordes, entre otros.

II.2.2. Segmentación

Uno de los pasos importantes del procesamiento de imágenes digitales en el campo de la visión artificial es la segmentación, que básicamente consiste en dividir una imagen en varias partes o segmentos de píxeles agrupados bajo cierto criterio mediante el uso de algún método de segmentación. El objetivo base de la segmentación es identificar y tratar regiones "significativas", las cuales posteriormente serán analizadas con fines tales como identificar un objeto o detectar los bordes pertenecientes a un objeto. Existen diversas técnicas o métodos de segmentación como puede ser el método del valor umbral,

segmentación basada en histogramas, segmentación basada en detección de bordes, entre otros. Algunos de ellos serán abordados brevemente a continuación.

Método del valor umbral o Image Thresholding

El método del valor umbral o conocido también como Image Thresholding en inglés es un conjunto de algoritmos que tienen como propósito realizar segmentación en imágenes en base al valor numérico que posee cada píxel dentro de una imagen, haciendo una comparación entre el valor del píxel contra un valor de umbral establecido. Se debe destacar que para utilizar el método del valor umbral se consideran los valores unidimensionales de los píxeles en la imagen a procesar (Gonzalez *et al.* (2002)) como por ejemplo el nivel de gris de cada píxel. Este algoritmo es considerado un método de segmentación orientado a píxeles ya que realiza las operaciones correspondientes comparando píxel por píxel.

Por ejemplo, se tiene una imagen en escala de grises (Figura 7) a la cuál se le aplica un proceso de thresholding pasando un valor de umbral, dando como resultado una imagen manteniendo exclusivamente los píxeles que se encuentren dentro de dicho umbral (Figura 8), las imágenes pueden ser encontradas en el siguiente vínculo (Wikipedia).

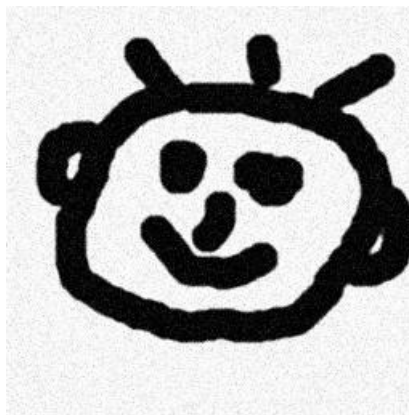


Figura 7. Imagen en escala de grises sin modificar



Figura 8. Imagen en escala de grises resultante de aplicar un valor de umbral

Segmentación basada en histogramas

Los métodos de segmentación que utilizan histogramas son, por lo general, muy eficientes en comparación con otros métodos, esto debido a que solo necesitan realizar un único recorrido por cada píxel. En estos métodos se calcula un histograma a partir de los píxeles que conforman la imagen, una vez que se tiene el histograma se utilizan los picos y los valles son usados para ubicar grupos en la imagen, se pueden usar cuestiones como el color o la intensidad como medida para calcular los grupos (Raju y Neelima (2012)). Una de las desventajas de utilizar la segmentación basada en histogramas es que puede resultar complicado identificar los picos y los valles que son realmente importantes en la imagen.

Segmentación basada en la detección de bordes

Existen técnicas de detección de bordes que pueden ser usadas como métodos de segmentación (Al-Amri *et al.* (2010)), esto debido a que por lo general los límites de los objetos o de las regiones de interés tienen un fuerte ajuste de intensidad, ya sea de color o de iluminación, por lo que es sencillo determinar y segmentar un objeto de

otro. Sin embargo, para poder realizar de forma correcta una segmentación basada en la detección de bordes es necesario que los bordes formen figuras cerradas concretas, ya que de otro modo se puede confundir un conjunto de bordes que daría como resultado un segmento único, lo cuál es un error cuando se trata de segmentar regiones aisladas de una imagen.

II.2.3. Detección de objetos

En la detección de objetos se trata de detectar casos de objetos semánticos de una cierta clase ya sea en una imagen digital o un video, esto con el fundamento de que cada objeto tiene sus propias características particulares que lo hacen diferente a las demás objetos que pueden o no estar presente en una imagen, por ejemplo, una persona tiene características particulares que lo diferencian de un vehículo o de un animal. La detección de objetos es una de las partes más importantes en la visión artificial, ya que de ello parten cosas como la clasificación de objetos dentro de una imagen, el seguimiento de objetos e inclusive la detección de características particulares, como la detección de rostros, detección de manos, detección de líneas, entre otros. Existen diversos métodos de detección de objetos, cada uno empleando una técnica particular, como puede ser el Template Matching, la detección de bordes, Feature Matching, entre otros. Los métodos previamente mencionados serán explicados brevemente a continuación.

Template Matching

El template matching es una técnica de detección de objetos que se basa en encontrar y emparar partes de una imagen en base a una "plantilla"; es decir, se toma la información procedente de la imagen que esta siendo analizada y se compara con la información de la imagen plantilla en la cual se tiene un patrón a identificar, por ejemplo

alguna forma o figura en particular, alguna similitud en la posición de los píxeles; en pocas palabras se comparan las imágenes en busca de similitudes partiendo de una imagen base para posteriormente examinar la cantidad de diferencias entre una y otra (Brunelli (2009)).

Detección de bordes

La detección de objetos mediante la detección de bordes tiene como fundamento principal que identificar puntos en una imagen en donde el brillo, el color o la intensidad de un objeto cambie de forma drástica o en otras palabras, que haya discontinuidades entre un objeto y otro (Papari y Petkov (2011)). Los cambios drásticos de brillo dentro de una imagen pueden significar sucesos importantes dentro del contexto de la imagen, por ejemplo, que haya discontinuidad en la profundidad, siendo así que un objeto puede encontrarse al frente de la escena mientras que se determina que lo que se encuentra detrás de dicho objeto es el fondo o es un objeto distinto detrás; también se puede determinar que hubo un cambio en las propiedades del material, un cambio en la iluminación, entre otros.

El uso de una detección de bordes puede resultar en líneas o curvas interconectadas que representan las fronteras de los objetos (Al-Amri *et al.* (2010)) y con esto se puede aligerar la información que será analizada y así descartar lo que se considere menos relevante, aunque en ocasiones la detección de bordes puede resultar contraproducente en objetos medianamente complejos, ya que se pueden crear bordes que no están realmente conectados entre sí (como se había mencionado en una sección previa), lo que produce un objeto complejo que pierde las características importantes y se vuelve difícil de analizar.

Feature Matching

Esta es una técnica de detección de objetos mediante la cual se busca extraer y emparar características específicas de los objetos presentes en la imagen, por ejemplo, alguna forma o figura en particular, algún color, textura, etc., de esta manera, si en la imagen objetivo se encuentra alguna similitud de esa característica particular se analiza para, posteriormente, realizar la detección del objeto.

El feature matching actualmente se está perfeccionando gracias a la implementación de redes neuronales y clasificadores de aprendizaje profundo. Esta técnica es ampliamente implementada en redes neuronales convolucionales, ya que la imagen va pasando entre cada capa de la red neuronal, las cuales van refinando la información extraída y así se obtiene información precisa que puede ser usada en la clasificación o detección de objetos.

II.3. Comportamiento animal

El comportamiento animal ha sido objeto de estudio incontables veces, esto con la finalidad de entender el por qué de las cosas y como es que los animales se adaptan a su entorno cambiante a través de los años, como es que nuestra interacción con ellos los ha llevado a adoptar diversas medidas ya sea para comunicarse, protegerse o esconderse de nosotros, además de intentar comprender cuestiones acerca de como el comportamiento afecta a la supervivencia y la reproducción de una especie, así como entender como el paso del tiempo los ha afectado a nivel evolutivo. Primero que nada, es necesario definir qué es comportamiento, que según [Dugatkin \(2020\)](#) la definición es "El comportamiento es el conjunto de respuestas coordinadas de todos los organismos vivos a estímulos internos y/o externos". Así pues, tomando en mente esta definición se

entrelazan y relacionan algunos términos, como la selección natural de Darwin, donde se argumenta en palabras resumidas que cualquier rasgo que pueda ser transmitido a través de cada generación y provea a un animal con algún tipo de ventaja reproductiva por sobre los otros miembros de la población en la que se encuentra será favorecido por la selección natural, y si bien la selección natural altera los comportamientos a través del curso de varias generaciones, el aprendizaje individual de cada organismo puede afectar la frecuencia con la que ciertos comportamientos se muestran a lo largo de su vida. Sin embargo el comportamiento que exhiben los animales también puede ser alterado en mayor o menor medida debido al aprendizaje social y la transmisión cultural (Dugatkin (2020)), entendiendo esto último como sistemas en los que los animales aprenden de su interacción directa con sus congéneres u otras especies en su entorno directo. Cada especie animal tiene un comportamiento específico ante las diversas situaciones que se le puedan presentar a lo largo de su vida, teniendo así sus rasgos característicos que definen tal o cuál actitud, y en este trabajo nos centraremos en el comportamiento canino.

II.3.1. Comportamiento canino y sus posturas

Antes de entrar de lleno en materia, es necesario definir lo que son los caninos, así pues es necesario entender que la subfamilia de los cánidos (Canidae) llamados caninos (Caninae) son aquellos mamíferos cuadrúpedos con aspectos de perro-lobo que tienen una dieta basada mayormente en carne (con algunas excepciones omnívoras), algunos de los ejemplos más claros que podemos encontrar en esta subfamilia son los lobos, perros salvajes y domésticos, coyotes, zorros, dingos, entre otros, los cuales también se dividen en grupos según sus variaciones filogenéticas (Wilson y Reeder (2005), Lindblad-Toh

et al. (2005), Wayne (1993)).

Ahora bien, los caninos tienen actitudes, gestos y comportamientos particulares que han sido ampliamente estudiados a lo largo de los años. En el presente trabajo de tesis se tomara como base el trabajo de Handelman (2012) que es una guía muy completa e ilustrada con fotos acerca de cada comportamiento que ha sido estudiado a la par de las posturas y acciones específicas de cada comportamiento, y es importante destacar que al ser un trabajo tan extenso solo se tomarán los conceptos clave necesarios para el desarrollo del presente documento, como es el comportamiento de la agresividad, la ansiedad (mencionado también como estrés) y el miedo.

Según Handelman (2012) "La agresividad es cualquier comportamiento que comunique un desafío, una amenaza o un intento de dañar a otro individuo incluyendo, pero no limitado a mostrarse como gruñidos, posturas o chasquidos". El comportamiento agresivo puede clasificarse a su vez en categorías específicas que refieren al detonante de dicho comportamiento, como por ejemplo comportamiento agresivo biológico, comportamiento agresivo competitivo, comportamiento agresivo defensivo, comportamiento agresivo por dominancia, entre otros.

Las principales posturas y acciones que acompañan a este comportamiento son por lo general, pero no limitado a (Handelman (2012)):

- mostrar los dientes.
- gruñir o ladrar.
- lanzar mordidas al agresor o rival.
- respiración fuerte con exhalaciones ruidosas.
- orejas en punta hacia el frente o hacia atrás acompañado de gruñidos o muestra

de los colmillos.

- cabeza agachada, cercana al nivel del suelo mientras que el cuerpo se encuentra medianamente agachado.

Entre otros.

Ahora en la cuestión de la ansiedad es importante señalar que se puede manifestar en todos los aspectos de la vida, en todas las situaciones, a cualquier edad, y así como pueden presentarse diversos comportamientos ante el estrés entre seres humanos pasa lo mismo entre caninos, es decir, es muy probable que al estar en una situación en específico, la reacción de un perro a otro se perciba de manera distinta, esto se puede entender como que a un perro puede no generarle ansiedad o estrés mientras que otro pueda sentirse claramente estresado por la misma situación y por lo tanto reaccionar diferente [Handelman \(2012\)](#).

Así pues, las principales posturas que se asocian a este comportamiento son las siguientes:

- Parpadeo más frecuente de lo usual.
- Lip licking (lamerse los labios/hocico).
- Rascado intenso.
- Sacudido intenso.
- Bostezo constante.
- Cauteloso / reacio a acercarse.
- Cola por debajo del nivel del lomo pero sin llegar a estar entre las patas.

Entre muchas otras dependiendo del nivel de estrés o ansiedad al que está sometido el perro.

Y por último, en lo relacionado al miedo se explica en [Handelman \(2012\)](#) que al tener miedo, por lo general el canino entra en modo reactivo", lo cuál inhibe hasta cierto punto la porción pensante de su cerebro. De este modo la mayoría de las acciones que el perro presente durante este comportamiento serán acciones evasivas o agresivas en orden de sortear/evitar la situación que le produce ese miedo.

Por lo tanto, algunas de las posturas más presentadas por los perros que tienen miedo son las siguientes:

- Esconderse u protegerse detrás de un objeto o de tratarse de un perro doméstico, detrás de su dueño.
- Esconder la cola entre las patas.
- Agachar la cabeza escondiéndola entre los omóplatos/escápulas.
- Temblores visibles por el cuerpo.
- Inclinar en la dirección contraria del objeto/animal/persona que le produce miedo
- Esconder las orejas hacia abajo y hacia atrás.
- Agresión desmedida en orden de alejar a lo que le produce miedo.

Por supuesto que estas no son las únicas posturas que puede presentar un perro al momento de tener miedo, esto dependerá del estímulo que lo provoca y de experiencias previas ante situaciones o estímulos similares.

Capítulo III

Estado del arte

Primero fue necesario realizar una investigación exhaustiva acerca de los trabajos que estuvieran relacionados con el problema a resolver para encontrar similitudes y diferencias, y con ello poder marcar la pauta de la metodología a diseñar comprendiendo lo que ya existe y que se puede mejorar en este campo, obteniendo la siguiente información.

III.1. Detección y clasificación de posturas mediante visión artificial

En el campo de la detección y clasificación de posturas mediante el uso de visión artificial existen diversos trabajos con sus métodos particulares y enfocados a detectar partes específicas, por ejemplo detectar posturas de manos mediante el uso de una red neuronal convolucional multicanal y un kernel cúbico ([Barros et al. \(2014\)](#)) donde se desarrolló un modelo capaz de analizar y clasificar las posturas de manos ingresando las imágenes procesadas de 3 formas distintas. Es decir, un canal de la red convolucional recibe las imágenes en escala de grises, otro canal recibe imágenes manipuladas mediante la aplicación de un filtro Sobel (cálculo de gradiente) en dirección "X" mientras que el último canal recibe imágenes después de aplicarse un filtro Sobel "Y", toda la información se concentraba en una última capa convolucional completamente conectada y obteniendo como resultado un sistema que permite realizar detección de manos en

tiempo real y mejorar los resultados contra los que se obtendrían usando un modelo regular con un kernel 2D.

Por otro lado en el trabajo de [Doan et al. \(2016\)](#) se crea un sistema robusto con la capacidad de reconocer posturas de manos en tiempo real utilizando imágenes RGB en conjunto con imágenes de profundidad provenientes de un sensor Kinect. Este trabajo se basa en la extracción de características mediante la segmentación de las regiones de la mano, utilizando un kernel donde se extraen tres niveles de características y se realizan tareas de segmentación, identificación y refinamiento del reconocimiento de manos usando diversos parámetros, primero sustrayendo el fondo de la imagen, posteriormente se realiza un enfoque y segmentación a priori en la región de interés que sería el centro de la mano, enseguida se hacen una serie de operaciones para calcular la altura máxima y el área que ocupa la mano y finalmente reconstruir la mano de manera digital y detectar la postura que está realizando.

Sin embargo, la detección de posturas no solo se centra en áreas pequeñas como la mano, sino que también se han realizado estudios para detectar posturas de cuerpo completo, un ejemplo de esto es el trabajo de [Elforaici et al. \(2018\)](#) en el que su enfoque constaba en realizar un sistema que sea capaz de detectar y clasificar posturas en humanos utilizando un dispositivo Kinect implementando redes neuronales convolucionales para un modelo de posturas en 2D y en 3D. En este trabajo se encargaron de utilizar imágenes RGB e imágenes de profundidad procedentes del dispositivo Kinect para entrenar un modelo que funcione con imágenes planas o en 2D, pero además implementar modelos del esqueleto humano en 3D de cada postura a clasificar, en este caso de las posturas De pie, Doblado, Sentado, Caminando, Agachado, más las imágenes de profundidad para hacer cálculos de ángulos geométricos y conjuntos de distancias para posteriormente clasificar y reconocer las posturas. Siguiendo con las

facilidades y las condiciones que brinda el uso del Kinect se han propuesto otros trabajos de reconocimiento de posturas como el de [Mealin *et al.* \(2016a\)](#) donde se propone utilizar tecnología que nos permita obtener mayor información visual, esto al utilizar la cámara RGB y la cámara infrarroja para analizar y clasificar mediante un modelo semi-supervisado las posturas de los perros, limitándose a clasificar posturas tales como estar de pie, sentarse y echarse o acostarse.

III.1.1. Visión artificial aplicada a animales

El uso de imágenes de profundidad o infrarrojas no siempre es necesario para detectar y clasificar posturas como lo demuestra el trabajo de [Nasirahmadi *et al.* \(2019\)](#) en donde el punto principal era determinar cual sistema de clasificación de imágenes bidimensional en combinación con técnicas de deep learning puede ser utilizado para detectar posturas en puercos que se encuentran en granjas comerciales.

Esto haciendo un análisis de métodos de detección basados en deep learning y redes neuronales convolucionales, enfocándose principalmente en las arquitecturas Faster R-CNN, SSD, R-FCN combinados con ResNet e Inception ResNet V2 para así poder realizar la clasificación e identificación de 3 posturas principales, De pie, Recostado sobre sus barrigas y Recostado de lado.

En este trabajo en particular solo fue necesario una cámara que proporcionaba imágenes RGB regulares, extrayendo un total de 4900 imágenes tomadas de forma aérea, las cuales fueron divididas en un set de entrenamiento, un set de validación y otro set para realizar pruebas, resultando en un sistema que se desempeña de manera correcta en la detección y clasificación de las posturas obteniendo los mejores resultados de clasificación a un ritmo de aprendizaje (Learning Rate) de 0.003 durante

el entrenamiento, obteniendo una precisión de 0.93 en la postura De pie, 0.92 Recostado de lado y 0.89 en la postura Recostado sobre su barriga.

En el campo de la detección de rostros mediante el uso de visión artificial, la detección de rostros humanos es un campo muy explorado y que actualmente existen una amplia gama de algoritmos robustos y funcionales para la detección de rostros humanos. Sin embargo, la detección de rostros animales es un campo poco explorado, por ello en el trabajo de [Mukai *et al.* \(2018\)](#) se crea un modelo capaz de realizar el reconocimiento facial de perros y gatos mediante la combinación de clasificadores construidos bajo técnicas diferentes, concretamente los autores entrenaron 4 clasificadores:

- Clasificador de rostros construido mediante un extractor de características Haar-like
- Clasificador de rostros construido mediante histogramas de gradientes orientados
- Clasificador de partes del rostro construido mediante un extractor de características Haar-like
- Clasificador de partes del rostro construido mediante histogramas de gradientes orientados

En este trabajo en particular los autores tenían como propósito crear un modelo clasificador híbrido a partir de la combinación de dos clasificadores entrenados de las maneras antes mencionadas, usaron un total de 1000 imágenes de gatos y 600 imágenes de perros de 3 razas distintas, 200 por cada raza.

Una vez que entrenaron y desarrollaron cada modelo los autores encontraron que la combinación del clasificador de rostros Haar-Like y el clasificador de rostros de histogramas de gradientes orientados resultaba en un modelo robusto y capaz de

detectar el rostro de los perros y los gatos con mayor precisión que cualquier modelo individual o que combinando cualquiera de los otros modelos, siendo capaz de detectar incluso un rostro reflejado sobre una superficie, obteniendo un valor de precisión del 75.7% en la detección de rostros de gatos y un valor de precisión de 90.8% en la detección de rostros de perros.

Detección de posturas en perros mediante visión artificial

Así como al inicio del capítulo se mencionaron algunos trabajos orientados en la detección de posturas en humanos, los trabajos de detección de posturas en perros no se han quedado atrás, como se puede observar en el trabajo de (Mealin *et al.* (2016a)) donde se propone diseñar un algoritmo semi-supervisado capaz de detectar posturas básicas como estar de pie o pararse, sentarse y acostarse utilizar la tecnología del dispositivo de Kinect, el cual posee una cámara infrarroja, una cámara de profundidad y una cámara regular; el procedimiento es bastante similar al utilizado en los trabajos mencionados al inicio de este capítulo, ya que haciendo uso de la cámara infrarroja y la cámara regular se capturan varios videos y se realiza un procesamiento de dichos videos para poder detectar las posturas en cuestión. Los autores del trabajo llevaron a cabo los siguientes pasos para realizar la detección de posturas:

1. Capturar múltiples imágenes de profundidad estáticas del background sin que el perro esté presente.
2. Conjuntar todas las imágenes previamente capturadas del background usando la media y el promedio de los pixeles vecinos en cada imagen para reducir el ruido producido por la cámara del Kinect.
3. Para cada imagen de profundidad con un perro sobre la escena, se sustrae el fondo

contemplando que todo lo que se encuentre a 20 milímetros o menos del suelo es parte del fondo o background.

4. Se utiliza detección de contornos para encontrar la localización exacta del perro en la escena.
5. Calcular la distancia promedio entre el perro y el Kinect usando la información extraída de la cámara de profundidad y la detección del contorno del perro.
6. Se capturaron las tres posturas a evaluar en intervalos de 10 segundos entre cada postura y posteriormente se etiquetaron de forma manual las imágenes resultantes de la grabación de dichas posturas.
7. Las imágenes etiquetadas se usaron para entrenar el algoritmo semi-supervisado y posteriormente se realiza su ejecución en tiempo real.

Bajo este método los mejores resultados obtenidos fueron de una precisión de 0.94 en la detección de la postura de estar de pie, un 0.91 en la postura de estar sentado y un 1.00 en la postura de estar echado o acostado.

Otro trabajo relacionado de los autores ([Mealin *et al.* \(2016b\)](#)) tuvo como finalidad reconstruir y analizar la posición de un perro en un escenario mediante un algoritmo no supervisado, utilizando una vez más el Kinect y haciendo uso de la cámara de profundidad y la cámara infrarroja, además de que al utilizar la información proveniente de la cámara de profundidad, la cámara infrarroja y una imagen binaria, los autores fueron capaces de reconstruir con mejor precisión la silueta o el cuerpo de un perro dentro de la escena, eliminar los elementos extras o innecesarios y con ello determinar la orientación del perro, es decir, si el perro se encontraba viendo de frente, si el perro estaba de costado, si estaba siendo visto desde la parte trasera del cuerpo o

si se encontraba inclinado en cierta dirección, para esto los autores realizaron tres experimentos bajo diversas condiciones y siguiendo tres metodologías distintas. Sin embargo cada metodología consistía en cambiar la posición de la cámara, cambiar el ángulo en que se encontraba el perro y por último segmentar las partes corporales del perro. Si se combina ambos trabajos de estos autores, sería posible crear un sistema capaz de detectar posturas básicas de perros con mayor precisión y además distinguir la orientación que tiene el perro durante dicha postura.

III.2. Estudios acerca del comportamiento canino

Por otro, lado existe una amplia cantidad de estudios acerca del comportamiento canino, sin embargo, no todos los estudios están estandarizados ni utilizan un lenguaje común que permita hacer investigaciones cruzadas entre todos los estudios realizados. En el trabajo de Jones y Gosling (Jones y Gosling (2005)) se realiza un análisis y un compendio de artículos y trabajos que se han desarrollado entorno al comportamiento canino entre 1934 y 2004, presentando los siguientes resultados:

- 43 estudios analizaron el miedo en los perros, relacionado mayormente con la reactividad a algún estímulo o impulso externo.
- 31 estudios hablaron sobre la sociabilidad en los perros, centrándose en el interés para entrenar a los perros, la interacción con otras personas y con otros perros.
- 34 estudios se centraron en estudiar la capacidad de respuesta al entrenamiento en los perros, enfocados en tendencias clasificadas como “Distracción” o “Concentración”, incluyendo “resolución de problemas, cooperación y disposición para trabajar”.

- Por otro lado, 30 estudios hablan sobre la agresividad de los perros, llevando a cabo pruebas un tanto invasivas como la aproximación de un extraño y posteriormente el ataque por parte del extraño hacia el perro o hacia el dueño del perro para registrar la reacción del animal. Cabe señalar que dependiendo de los autores de las investigaciones existían subcategorías en la agresividad.
- 16 estudios se enfocaron en la dominancia y la sumisión que podía presentar un perro ante diferentes escenarios, por ejemplo, no moverse del camino de una persona.
- 15 estudios estaban orientados a analizar la conducta de la actividad, es decir, los movimientos que realiza un perro, su coordinación locomotora, entre otros.
- Además, los autores Jones y Gosling encontraron 23 artículos que no eran miembros de las categorías anteriores pero que también estudiaban el comportamiento de los perros.

III.2.1. Estudios acerca del comportamiento canino con aplicación de modelos computacionales y visión artificial

Las posturas de los perros no ha sido el único campo estudiado mediante modelos computacionales, sino también su comportamiento.

En el trabajo de (Robinson *et al.* (2015)) se menciona que durante alguna emergencia o en alguna situación inesperada, los perros tienden a presentar acciones que, bajo condiciones normales no mostraría, como por ejemplo, puede que empiecen a correr, derriben a las personas que se encuentren cerca o por el contrario no actúen o ni realicen alguna acción ante la emergencia por lo que los autores mencionan que monitorizando

los patrones conductuales de los perros durante una situación de emergencia se pueden desarrollar tecnología capaz de pedir ayuda durante dicha situación inesperada.

Los autores proponen que una solución puede ser captar los signos vitales y de movimiento del perro mediante sensores colocados sobre el perro, ya que los perros pueden reaccionar mediante posturas o movimiento, además también mediante vocalizaciones irregulares que indiquen estrés o una afección negativa, también se menciona que si estos datos biométricos se combinan con cámaras que indiquen su posición dentro de un área determinada, un sistema puede enviar una señal de auxilio a los servicios de emergencia para que arriben a dicha ubicación o se sepa que en esa área específica sucedió algo inesperado que requiere atención humana.

Los autores mencionan algunos trabajos más que están desarrollando otras soluciones al problema y como se pueden usar ciertas tecnologías para complementar dichos sistemas.

Por otro lado, estimar o predecir los movimientos y comportamientos que puede tener un perro ante ciertas situaciones es lo que se hizo en el trabajo de (Ehsani *et al.* (2018)) donde presentan un sistema computacional en el que a partir de sensores y cámaras acopladas a un perro se modela el comportamiento de un perro, extrapolando los movimientos registrados a un modelo 3D utilizando técnicas de digitalización para estudiar el posible comportamiento y las acciones que podría llevar a cabo un perro a la hora de realizar ciertas actividades; este análisis se realiza mediante un modelo computacional que analiza imágenes capturadas en video y recrea los movimientos probables que llevará a cabo el perro. Estas imágenes son grabadas con una cámara GoPro montada sobre el perro a la altura de la cabeza, además se recopila información del movimiento y posición corporal del perro durante el trayecto mediante unidades de medida inercial o IMUs por sus siglas en inglés (Inertial Measurement Units) cuatro

IMUs recopilan información de las extremidades, un IMU mide la actividad de la cola y una más se encarga de medir la posición del cuerpo, las IMUs permiten capturar el movimiento del perro en términos de desplazamiento angular.

Estos datos se analizan posteriormente por una red neuronal convolutiva ResNet, la información se codifica y se envía al modelo computacional encargado de recrear los movimientos que el perro realizó durante el trayecto. Una vez que el modelo se entrenó lo suficiente, el sistema es capaz de predecir con cierto grado de exactitud los movimientos futuros que podría realizar un perro dadas ciertas condiciones o ante ciertos estímulos.

Así pues, como se puede observar existen una amplia variedad de trabajos que hacen uso de modelos computacionales y visión artificial para resolver problemas de clasificación de posturas tanto en humanos como en animales, cada uno desde sus enfoques particulares pero atacando una misma problemática compartida. Además ahora existen trabajos dedicados a la detección de rostros entre distintas especies de animales, ya que si bien la detección de rostros en humanos es un campo ampliamente investigado y aplicado, los rostros de los animales tienen características particulares entre cada especie con lo cuál es necesario realizar trabajos e investigaciones exhaustivas enfocadas a cada especie para crear un modelo de detección de rostros extrayendo esas particularidades que no se comparten entre especies.

Y por el lado del comportamiento canino si bien existen algunos trabajos que atacan ciertas problemáticas particulares de perros domésticos, hace falta más investigación y desarrollo de nuevas tecnologías para clasificar y detectar comportamientos y actitudes comunes del perro relacionados a su estado interno, su forma de ser y de actuar ante ciertas situaciones particulares. Dicho esto, una vez que se analizaron las propuestas y las soluciones de los trabajos previamente descritos este proyecto intenta incursionar en el campo de la clasificación del comportamiento del perro doméstico, tomando algunas

de las técnicas utilizadas para la detección de posturas o rostros y añadiendo los estudios acerca del comportamiento canino, además de sumar algunas otras técnicas de visión artificial y redes neuronales convolutivas, dicha metodología es descrita a detalle en el capítulo **IV**.

Capítulo IV

Metodología

La mayoría de los trabajos que hacen uso de visión artificial aplicado a perros se enfocan principalmente en la detección de posturas y detección de rostros, además unos pocos se enfocan en generar soluciones a diversas problemáticas relacionadas con el comportamiento de los caninos, sin embargo el campo de la detección y clasificación del comportamiento haciendo uso de modelos computacionales y visión artificial aún sigue siendo un campo poco explorado, por lo que en este capítulo se propone una metodología para llevar a cabo la clasificación y detección de comportamientos del perro doméstico haciendo uso de las tecnologías más recientes en visión artificial y redes neuronales convolutivas.

Dicho esto, en esta tesis se propone un método para la detección y clasificación del comportamiento en los perros, más específicamente para detectar el comportamiento de la agresividad, la ansiedad y el miedo, tomando a consideración las acciones y movimientos más visuales que presentan los perros, esto con la finalidad de proveer información complementaria en el proceso de selección y entrenamiento de perros de rescate. En este capítulo se describe los pasos y procesos que se llevaron a cabo durante el desarrollo de este trabajo.

IV.1. Corpus de datos

El presente trabajo utiliza el corpus de datos proveniente de videos recopilados durante otro proyecto del Doctor Humberto Pérez Espinosa en el año 2015 y 2017,

estos datos ahora pertenecen a la empresa Mescalina quienes permitieron su uso con fines de investigación y desarrollo.

El corpus de datos está compuesto por 386 videos de 121 perros distintos, cada video es de una duración variable, algunos pueden durar desde unos pocos segundos hasta horas. Además, fueron grabados en ambientes no controlados, siendo mayormente los domicilios particulares de los dueños de los perros, por lo que la cantidad de objetos en escena, movimiento de la cámara, el color del fondo, la iluminación del lugar, entre otros factores que llegan a ser muy variables. Esto puede afectar positiva o negativamente al procesamiento y desarrollo. Debido a lo anteriormente descrito es normal que no todos los videos en cuestión puedan ser usados debido a la dificultad para procesarlos. Por lo cual fue necesario realizar un pre-procesamiento para filtrar los videos que no fueran útiles, ya sea debido a que no contenían información relevante o que la visibilidad fuera nula. A continuación se explica como se llevó a cabo este pre-procesamiento.

En adición a los videos, se utilizó una base de datos pública que lleva por nombre Dogs vs. Cats ([Kaggle, 2014](#)), la cual contiene 25000 imágenes en total entre las categorías de perros y gatos, este conjunto de datos sirvió como apoyo para tener aún más imágenes de perros para el entrenamiento del modelo.

La figura [9](#) muestra de manera resumida el esquema general de los pasos realizados durante cada fase del proyecto, y a continuación se describen detalladamente las actividades que conformaron cada paso en su respectiva sección.

IV.2. Pre-procesamiento del corpus

En esta fase fue necesario realizar una revisión manual de cada video para identificar cuales videos contenían información útil y cuales podían descartarse. El proceso de

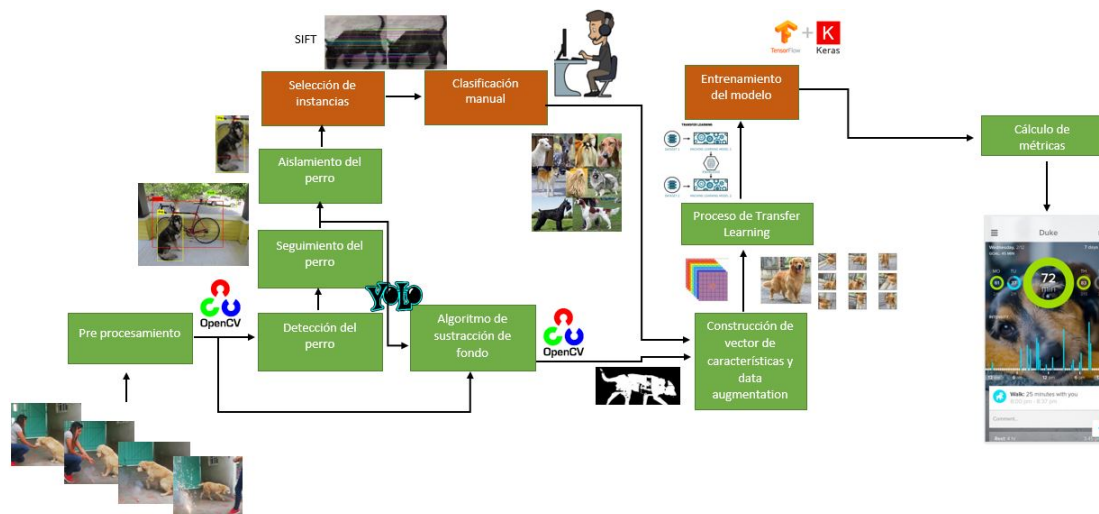


Figura 9. Esquema general de actividades

revisión manual consistió en lo siguiente:

1. Se realizaron dos tablas, una para los videos del año 2015 y otra para los videos del año 2017 en una hoja de cálculo. En dichas tablas los nombres de cada perro conformaban las filas, mientras que las columnas pertenecían a los comportamientos, es importante señalar que la cantidad de videos por cada perro podía variar, es decir, un perro podía tener un único video mientras que otro podía tener 2 o más.
2. Una vez estructuradas las tablas se procedía a reproducir uno por uno cada video de cada perro, es decir, se revisaron manualmente todos los videos pertenecientes a cada perro con el fin de evaluar la utilidad de cada uno.
3. Para realizar la evaluación de los videos se veía la totalidad del video y se evaluaban cosas tales como la visibilidad de los perros, las acciones de los perros y en que comportamiento podría incurrir cada acción, además de cuestiones como la iluminación de la escena, el movimiento y enfoque de la cámara. Por ejemplo,

si en un video no había suficiente iluminación y el perro no se ve, ese video queda descartado, al igual que si hay demasiados obstáculos delante del perro o no se aprecia claramente los gestos y movimientos del perro debido al movimiento de la cámara. Por otro lado, si el perro permanecía sin realizar alguna acción durante todo el video aún si la visibilidad era perfecta también se descartaba.

Para catalogar las acciones se cotejaba lo que se observaba con una lista de las acciones/posturas más representativas pertenecientes a los comportamientos de agresividad, ansiedad y miedo. Esta lista nos fue provista por la Dra. Verónica Reyes del Centro Tlaxcala de Biología de la Conducta quien es experta en el análisis del comportamiento animal.

4. Una vez que se identificaba una acción perteneciente a algún comportamiento y se hubiese comprobado que los demás elementos como la visibilidad o la iluminación no fuesen un problema, se anotaba el nombre del video en la columna correspondiente al comportamiento identificado. Si el nombre consistía únicamente en dígitos se anotaban los últimos 3 o 4 dígitos del nombre, si el nombre era alfanumérico se anotaba tal cual. Si el video era descartado por falta de acciones por parte del perro el renglón con su nombre se dejaba en blanco, por otro lado si el descarte era debido a los elementos como la visibilidad, iluminación, obstáculos, etc., se realizaba una anotación señalando el motivo de dicho descarte.

Es importante señalar que un mismo video podía ser catalogado en más de una categoría de comportamiento, es decir, si durante el video el perro se mostraba agresivo durante algunos minutos y después se mostraba ansioso o con miedo, el nombre del video se anotaba en las columnas correspondientes.

La cantidad de videos clasificados en cada categoría se muestra en la tabla [II](#).

Tabla I. Número de videos en cada categoría de acuerdo a la revisión y clasificación manual.

Año	Agresividad	Ansiedad	Miedo
2015	57	44	12
2017	128	72	15
Total	185	116	27

IV.3. Generación de muestras

Una vez que se listaron y catalogaron los videos de manera manual, el siguiente paso consistió en la codificación de un programa que fuese capaz de detectar de forma automática a los perros dentro de los videos mencionados en la sección anterior, y que una vez los haya detectado realice un recorte automático del(los) perro(s) detectado(s) enfocándose o encuadrando exclusivamente al perro eliminando el resto de la escena lo mejor posible. El programa está codificado/diseñado en el lenguaje de programación Python haciendo uso de la librería OpenCV con herramientas para visión artificial y fue implementado sobre la plataforma de Google Colaboratory dadas las facilidades que ofrece dicha plataforma. Este programa tuvo varias etapas y evoluciones como se describe en las sub-secciones siguientes.

IV.3.1. Detección y seguimiento del perro

La primera fase consistió en lograr que el programa fuese capaz de detectar y realizar seguimiento de perros en tiempo real dentro de la escena del video, esto se logró auxiliandose del estado del arte en visión artificial llamado YOLOv3 (Redmon y Farhadi (2018)) que es un sistema de detección de objetos en tiempo real basado en la red neuronal de código abierto DARKNET (Redmon (2013–2016)), YOLOv3 sirvió como el sistema de detección de objetos base gracias a su flexibilidad para adaptarlo a

clases personalizadas así como alta precisión para detectar y seguir objetos (Redmon y Farhadi (2018)); se utilizó una versión pre-entrenada de YOLOv3 que además de detectar perros era capaz de identificar otros objetos, sin embargo el código se adaptó de forma que solo se extraía la información relacionada a la detección de perros. Los resultados se pueden observar en las figuras 10 y 11, donde en la figura 10 se observa la detección general de los objetos en la escena y posteriormente en la figura 11 se puede observar el aislamiento en la detección enfocando únicamente a un perro.

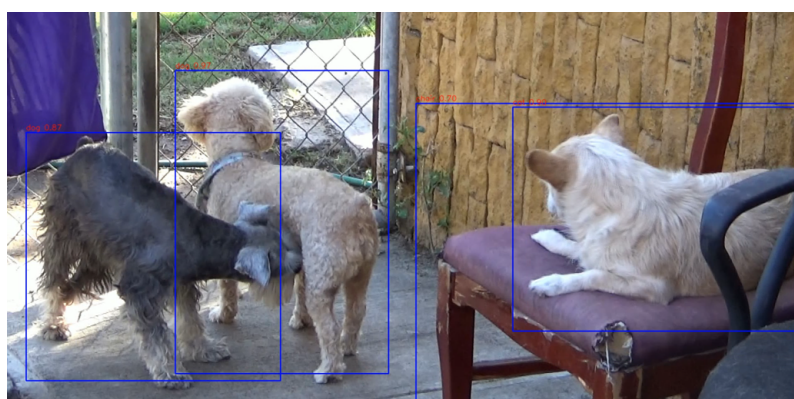


Figura 10. Detección general de objetos en escena



Figura 11. Aislamiento en la detección enfocando únicamente al perro

Una vez que se logró esta primer etapa, lo siguiente fue lograr que el sistema realizara recortes y almacenamientos automáticos de los perros detectados ignorando el resto de objetos en la escena.

IV.3.2. Aislamiento del perro y recorte automático

Lo siguiente fue lograr que el programa fuese capaz de aislar, recortar y almacenar de manera automática los perros que se detectasen en cada frame del video; esto dado que la cantidad de frames en un video podía llegar a ser sumamente extensa para realizar recortes manuales (por ejemplo, un video con una duración de 3 minutos y medio podía contener en promedio 7000+ frames), además de que no en todos los frames se podía apreciar visualmente a los perros por lo que algunos frames resultaban carentes de valor como material para la base de datos. Haciendo uso de la información proporcionada por la detección y aislamiento obtenido en la fase inicial se logró realizar que el programa fuese capaz de realizar la tarea del recorte y almacenamiento automático, inicialmente almacenando todas las imágenes recortadas de los frames en donde se hiciese una detección de un perro. Para que el almacenamiento de cada recorte se realizara de manera automática los nombres de las imágenes almacenadas se limitaban a números consecutivos iniciando por el número uno como se puede observar en la figura [12](#).

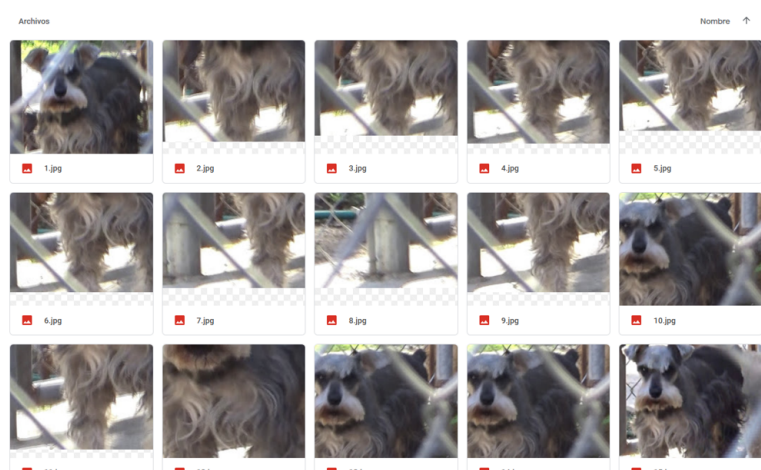


Figura 12. Recortes almacenados con numeración consecutiva

Sin embargo, la cantidad de imágenes similares entre sí era muy alta, es decir, podían

existir 10 o más recortes donde el cambio en la postura o la acción del perro era mínima y los recortes eran iguales o al menos similares, esto debido al tiempo transcurrido dentro del video o que simplemente el perro no se movía ni realizaba ninguna otra acción durante un periodo de tiempo, por lo que el siguiente paso fue lograr que el programa descartase los recortes que fuesen sumamente similares entre sí.

IV.3.3. Selección de instancias

En esta etapa lo que se buscaba lograr era que el programa hiciese una comparación entre recortes y si los recortes comparados eran similares, uno de los recortes se descartara, por otro lado si los recortes eran diferentes se almacenaran. Para lograr dicha tarea se recurrió a implementar un algoritmo de comparación SIFT (Lowe (2004)) logrando así reducir la cantidad de imágenes similares almacenadas.

El algoritmo SIFT (Scale Invariant Feature Transforms) es uno de los métodos de extracción de puntos invariantes existentes, su nombre se debe a que realiza transformaciones de los datos de una imagen en coordenadas invariantes a cambios como la escala, la rotación, la iluminación (hasta cierto punto) y al punto de vista en 3D si se implementa en un problema con estas características (Lowe (1999)). Los datos extraídos se consideran una característica de la imagen, dicha característica es un punto de interés único tomando en cuenta su posición, escala, orientación y su vector descriptivo. Así pues, durante las etapas iniciales del proyecto era necesario implementar un algoritmo que mediante los descriptores que arrojase fuese posible comparar dos imágenes y determinar que tan parecidas eran entre sí. Debido a las características de las imágenes con las que se trabajó era necesario que el algoritmo a implementar fuese capaz de detectar dos puntos de interés que fuesen idénticos a pesar de un cambio

ligero en la iluminación, las dimensiones relativas entre las imágenes, o si la imagen había rotado un poco debido al movimiento de la cámara con la que se grabaron los videos, eso sin contar el ruido que puede existir en las imágenes extraídas de un video originado por el mismo movimiento.

El proceso para comparar los recortes es el siguiente:

1. Se toman dos recortes, siendo en primer lugar el recorte almacenado más reciente, es decir, el último que fue almacenado por el sistema (de ahora en adelante referido como recorte 'A') y en segundo lugar el recorte que se acaba de obtener mediante la detección automática durante la ejecución del programa (de ahora en adelante referido como recorte 'B').
2. Se verifica las dimensiones de ambos recortes, de ser iguales en dimensiones ambos recortes pasan a la siguiente etapa, de lo contrario ambos recortes se reajustan a las mismas dimensiones.
3. Una vez que ambos recortes tienen las mismas dimensiones se procede a su comparación mediante el algoritmo SIFT como se observa en la figura 11. Si el porcentaje de coincidencia entre ambas imágenes es superior o igual al 40 % el recorte 'B' (es decir, el que se acababa de obtener durante la ejecución del programa) se descarta, ya que se considera que es muy similar al recorte 'A' y se considera innecesario almacenar dos recortes tan similares, por lo que el recorte 'A' se mantiene para la siguiente comparación.

Sin embargo, si el recorte 'B' se mantiene por debajo del 40 % se considera que tiene una diferencia significativa con respecto al recorte 'A' por lo que se almacena y ahora el recorte 'B' se usa para la siguiente comparación, es decir, al

ser el último recorte almacenado este pasa a convertirse en el nuevo recorte 'A' y el siguiente recorte generado por el sistema pasa a ser el nuevo recorte 'B'.

Además, mediante operaciones basadas en el tiempo transcurrido entre cada frame y por lo tanto el tiempo relativo entre cada recorte comparado se logra reducir la cantidad de imágenes almacenadas sin perder aquellas que pueden contener información relevante para el proyecto.

IV.3.4. Algoritmo de sustracción de fondo

Cuando se trata de detectar posturas, algunos trabajos han optado por extraer el fondo de la escena y trabajar con la silueta resultante (Barros *et al.* (2014), Doan *et al.* (2016), Mealin *et al.* (2016b), Elforaici *et al.* (2018)) esto bajo el principio de que la silueta resultante contiene la información necesaria para distinguir entre una postura y otra. Para poder extraer únicamente la silueta es necesario separar el Foreground (frente o primer plano) y el Background (fondo o segundo plano) y para ello existen los algoritmos de sustracción de fondo, estos algoritmos se aplican a un video o una sucesión de imágenes que formen parte de la misma secuencia, en donde el fondo cambia muy poco o nada y en cambio, en el primer plano hay objetos moviéndose a lo largo de la escena constantemente. Tomando en cuenta los resultados de los trabajos antes mencionados se decidió implementar un algoritmo sustractor de fondo MOG a la par que se realizaba la detección y comparación regular, así pues se obtienen imágenes RGB convencionales e imágenes de las siluetas del perro en blanco y negro los cuales están numerados a la par que las imágenes RGB convencionales. Es decir, la imagen 50 obtenida mediante la sustracción de fondo corresponde a la imagen 50 en RGB y así sucesivamente.

IV.4. Clasificación manual

Una vez que el programa fue capaz de generar las muestras de manera automática, el siguiente paso consistió en realizar la clasificación manual de dichas muestras en las categorías con las que se trabajaría durante el proyecto las cuales son:

- Agresividad
- Ansiedad
- Miedo
- Nada

El proceso de clasificación se llevó a cabo revisando manualmente cada una de las imágenes generadas a partir de cada video, donde cada imagen se revisaba visualmente y cada detalle se cotejaba con una lista de las acciones visuales más comunes que podían indicar que un perro está mostrando un comportamiento específico. Esta definición de acciones o conductas nos fue proporcionada por la Dra. Verónica Reyes, investigadora del Centro Tlaxcala de Biología de la Conducta, la tabla de las conductas se puede ver en la tabla [III](#).

Se decidió agregar la categoría 'Nada' para incluir en ella las posturas o conductas que no estén asociadas a ninguna de las otras categorías. Esto con la final de evitar falsos positivos por parte del sistema de clasificación.

Tabla II. Lista de conductas a observar

Categoría	Conducta	Definición
Agresividad	Mostrar dientes	Retrae los labios y muestra los dientes apretados, generalmente se acompaña de gruñido
	Erizar pelo del lomo	Levanta el pelo de la parte más anterior del lomo
	Gruñir	Vocalización de /r/ emitida de forma constante, se considera frecuencia y duración
	Orejas hacia adelante	Inclinadas o elevadas ligeramente curvadas hacia adelante
	Ladrido	Se considera frecuencia y duración de cada uno
	Lanzarse contra las barras	Intentando romper o estrellar las barras de la jaula
	Exceso de salivación	Produce saliva que puede observarse alrededor de su boca
	Elevación de cola	La cola está en posición vertical por encima de la altura del lomo, sin movimientos oscilatorios
Ansiedad/ Nerviosismo/ Intranquilidad	Aullido	Vocalización aguda asociada a malestar
	Rascado intenso	El perro se rasca intensamente con alguna pata trasera. En ocasiones presenta ausencia de pelo o lesión producto del rascado
	Mordisqueo	Descartar infección o presencia de pulgas y otros parásitos
	Liplicking	Lengua ligeramente fuera de la boca recorre hocico, labios y nariz
	Elevación ligera de cola	La cola no cae verticalmente, está elevada pero sin pasar la altura del lomo del perro
	Hiperactividad	Movimiento excesivo que incluye levantarse, echarse y caminar en círculos dentro de la jaula
	Acicalamiento excesivo	Se lame o mordisquea intensamente diferentes partes del cuerpo
	Lamer objetos persistentemente	Puede ser su bebedero, comedero o alguna parte de su jaula
	Sacudir el cuerpo	Similar a lo que hacen después del baño para retirar el agua de su cuerpo
	Bostezo	En horarios en que habitualmente no duerme
	Lamer costillas	Pasar la lengua sobre sus costillas de forma persistente sin presentar lesión
	Resoplar	Expulsar el aire con fuerza por la nariz produciendo un sonido
	Miedo	Retracción del cuello
Postura baja		Patas delanteras completamente flexionadas, patas traseras flexionadas ligeramente Orejas hacia atrás y cola entre las patas
Cola escondida		La cola se encuentra flexionada bajo el vientre del perro
Orejas hacia abajo		Inclinadas hacia abajo y atrás
Aullar		Considerar duración y frecuencia
Rascar la jaula		Específicamente la puerta de entrada
Quejarse		Sonidos que no llegan a ser aullido pero muestran incomodidad
Curling up		El perro se enrosca, similar a cuando tiene frío
Temblor		Se aprecian vibraciones en su cuerpo en respuesta a un estímulo

IV.5. Construcción de vector de características

Una vez que se obtuvieron y clasificaron todas las imágenes en sus respectivas categorías, el paso previo al entrenamiento del modelo consistió en generar el vector de características aplicando técnicas de Data Augmentation que consiste en generar muestras sintéticas a partir de rotaciones, sesgos, enfoque y desenfoque parcial o total sobre un área de la imagen, efecto espejo, entre otras técnicas (Zoph *et al.* (2020)).

Esto con la finalidad de tener más muestras durante el entrenamiento, posteriormente se hace la transformación de las imágenes a vectores numéricos que contenían la información de cada color, el número de canales de cada imagen, dimensiones ajustadas para todas las imágenes, etiquetas de clases y también se aplicaron operaciones de

normalización sobre cada vector, teniendo así vectores normalizados por cada categoría y por imagen.

Una vez que los vectores de características han sido normalizados y preprocesados, se procedió a realizar el entrenamiento del modelo.

IV.6. Entrenamiento del modelo

Para el entrenamiento del modelo se decidió utilizar Transfer Learning (Torrey y Shavlik (2010)) para aprovechar las propiedades de las arquitecturas de redes neuronales robustas probadas y pre-entrenadas sobre categorías similares, se entrenaron varios modelos con base en distintas arquitecturas para comparar y elegir el modelo que mejor cumpla el propósito de este proyecto. Uno de los modelos se entrenó bajo la arquitectura de ResNet50 (He *et al.* (2016)), cargando los pesos de ImageNet (Deng *et al.* (2009)) y utilizando Transfer Learning con una red neuronal basada en esta arquitectura que había sido utilizada para clasificar entre dos categorías: Perros y gatos, tomando esa base se utilizó la arquitectura pre-entrenada de ResNet50 como extractor de características, se congelaron algunas capas y se eliminaron las capas finales que se encontraban totalmente conectadas y fueron sustituidas por las capas totalmente conectadas personalizadas dedicadas a realizar la clasificación requerida para el proyecto. Posteriormente se realizó un proceso de Fine Tunning sobre el modelo ya entrenado con el conjunto de datos específico de este proyecto.

Otro de los modelos entrenados para el proyecto se creó utilizando Transfer Learning bajo la arquitectura VGG-16 (Simonyan y Zisserman (2014)) usando bottleneck features pre-calculados, estas características o features son una representación abstracta de las imágenes luego de haber pasado por las capas inmutables (congeladas) de la

red neuronal. Dichos bottleneck features provenían de un clasificador de razas de perros, clasificando en total 133 razas de perros ([Página de DataSmart](#)), el proceso de adaptación específica para el proyecto fue similar al realizado sobre la arquitectura ResNet50, se congelaron la mayoría de las capas para ser utilizadas como extractores de características, se eliminaron las capas finales y se sustituyeron por capas totalmente conectadas para realizar la tarea de clasificación particular del proyecto.

Por último, se realizó otro modelo utilizando la arquitectura MobileNet, que es un modelo diseñado para ser utilizado en aplicaciones móviles, y es el primer modelo de visión por ordenador móvil de TensorFlow [Howard *et al.* \(2017\)](#), este modelo utiliza convoluciones separables en profundidad, lo que reduce significativamente el número de parámetros en comparación con la red con convoluciones regulares con la misma profundidad en las redes. Este proceso da lugar a redes neuronales profundas y ligeras. La convolución separable por profundidad consta de dos operaciones:

- La convolución en profundidad, que es un mapa de una sola convolución en cada canal de entrada por separado. Por lo tanto, su número de canales de salida es el mismo que el número de los canales de entrada.
- Convolución puntual, que puede explicarse como una convolución con un tamaño de kernel de 1x1 que combina las características creadas por la convolución en profundidad.

Esta arquitectura de MobileNet tiene una precisión del 70.6% en el conjunto de ImageNet en comparación con GoogleNet y VGG 16, que obtuvieron un 69.8% y 71% respectivamente [Howard *et al.* \(2017\)](#). Además, existen otros modelos más finos que mejoran aún más la arquitectura de MobileNet llamados Thin MobileNet [Sinha y El-Sharkawy \(2019\)](#). Sin embargo, para este trabajo, hemos utilizado la

arquitectura estándar de MobileNet bajo una herramienta llamada Teachable Machine (Máquina Enseñable) que entrena un clasificador en Tensorflow con las clases e imágenes proporcionadas por el usuario, la herramienta permite modificar parámetros específicos, pero la mayoría de ellos ya están predeterminados. Esta herramienta está desarrollada en Tensorflow.js, Javascript, y construida sobre MobileNet y pre-entrenada en 1000 categorías sobre ImageNet (Carney *et al.* (2020)); este proyecto también forma parte de Experiments with Google (Google Creative Lab, Teachable Machine) y es un proyecto de código abierto alojado en GitHub al que se puede acceder directamente desde su página web. La ventaja de utilizar dicha herramienta es que te permite añadir tantas muestras y clases como necesites sin necesidad de invertir demasiado tiempo modificando la arquitectura, cabe mencionar que dicho proceso si se tuvo que realizar en los otros modelos previamente descritos durante la fase de preparación del modelo; volviendo a Teachable Machine al finalizar el entrenamiento te permite probar el modelo en tiempo real con muestras obtenidas mediante la cámara web o mediante archivos locales, y por último te permite descargar el modelo entrenado en formato TensorFlow + Keras, TensorFlow Lite, TensorFlow JavaScript o alojarlo en el sitio web para compartir el enlace y que otros usuarios puedan probar el modelo.

IV.7. Cálculo de métricas

Por último se realizó un prototipo de aplicación web en donde se muestra el modelo en acción a través de una cámara web a la que se le muestran imágenes impresas. El modelo hace las predicciones en tiempo real, mostrando en pantalla la probabilidad de que una imagen pertenezca a una de las 4 clases ya mencionadas en una escala de valores discretos del 0.0 al 1.0.

El prototipo de de la aplicación se puede ver más a detalle en el capítulo [VI](#).

Capítulo V

Resultados experimentales

En este capítulo se presentan los resultados obtenidos durante la implementación de los algoritmos de comparación entre imágenes y el algoritmo de sustracción de fondo, la experimentación en el entrenamiento de los modelos de clasificación y la selección de arquitecturas.

V.1. Algoritmo SIFT para la comparación y descarte de imágenes

El algoritmo SIFT fue de gran ayuda para extraer estas características y permitir la comparación entre dos imágenes dado que cubría los aspectos fundamentales requeridos para realizar la comparación, además de que gracias a la robustez del algoritmo, sin embargo también se consideraron otros algoritmos como la comparación de histogramas y comparación de plantillas (Match Template), pero dado que de una imagen a otra podía cambiar 'mucho' el escenario debido únicamente a pequeños detalles en la imagen como la iluminación o el movimiento de la cámara, la comparación por histogramas y de plantillas se descartaron. Además de que SIFT es un algoritmo robusto que obtuvo buenos resultados durante la experimentación, en las figuras [13](#) y [14](#) se pueden observar dos comparaciones mediante el algoritmo SIFT entre dos imágenes diferentes y dos idénticas respectivamente.

Las imágenes no tienen el mismo tamaño, serán reajustadas automáticamente
 Puntos clave de la 1ra imagen 547
 puntos clave de la 2da imagen 100

Porcentaje de coincidencia 6.0 %



Figura 13. Comparación SIFT de imágenes diferentes

Las imágenes son completamente iguales
 Puntos clave de la 1ra imagen 61
 puntos clave de la 2da imagen 61
 Porcentaje de coincidencia 100.0 %

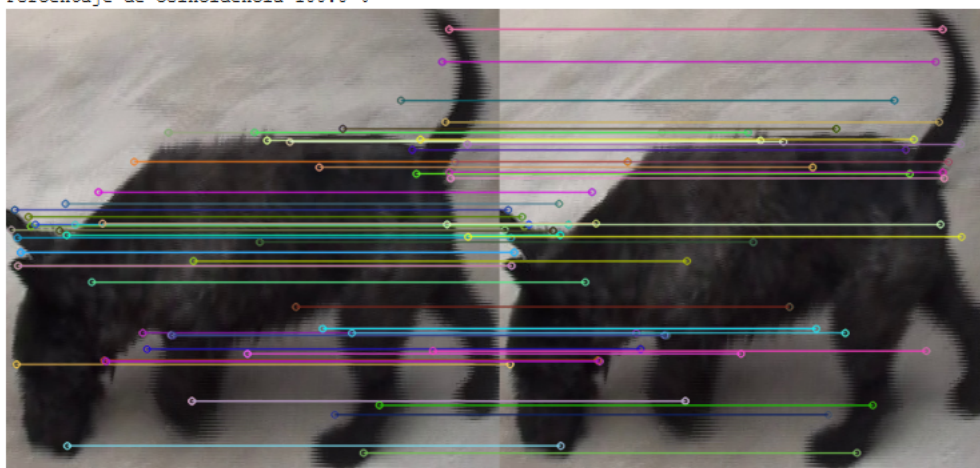


Figura 14. Comparación SIFT de imágenes iguales

La implementación del algoritmo redujo en gran medida las muestras que tenían un alto grado de similitud, como por ejemplo, se llevo a cabo un análisis de un mismo video con y sin la implementación del algoritmo SIFT y se extrajeron las muestras resultantes, tras dicho análisis se obtuvieron 5873 imágenes vs 16596 imágenes respectivamente,

esto sucede así ya que las muestras se obtienen frame por frame, lo que resulta en que en ocasiones existan muchos frames casi idénticos debido a que 1 frame es apenas una fracción de segundo dentro del video, por lo que suponiendo que la velocidad de captura de la cámara en cuestión es de 30 FPS (Frames Per Second) 30 frames representan apenas un segundo de avance, y si no hay ningún tipo de movimiento o cambio relevante durante dicho segundo se crean 30 muestras idénticas entre sí, lo cuál resulta poco útil.

V.2. Algoritmos de sustracción de fondo

Se decidió implementar un algoritmo de sustracción de fondo para obtener las siluetas de los perros a la par que se realizaba el seguimiento en y recorte en RGB. Se implementó el algoritmo MOG que estaba incluido en las librerías de OpenCv, un ejemplo del resultado se puede observar en la figura [15](#), como se puede ver el algoritmo es capaz de extraer la silueta del perro aún con algunos objetos de por medio, sin embargo es importante destacar que el algoritmo en ocasiones tiene problemas para realizar una detección clara del Foreground y del Background debido a factores como el movimiento de la cámara, cambios bruscos en el fondo o cambios en la posición desde la que se graba el video, esto se puede observar en las figuras [16](#) y [17](#).

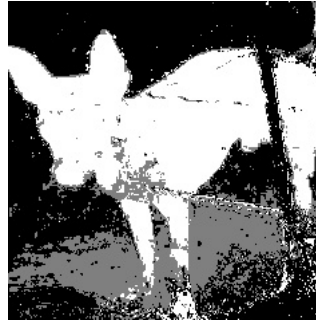


Figura 15. Silueta extraída con el algoritmo de sustracción de fondo



Figura 16. Problemas para extraer correctamente la silueta del perro después de un movimiento brusco de la cámara



Figura 17. Sustracción del fondo errónea debido a un cambio repentino en el fondo por un reposicionamiento de la cámara

Cuando la cámara se encuentra en constante movimiento, cambiando de enfoque,

de dirección o de escena es un tanto complicado llevar a cabo la sustracción del fondo de forma limpia y eficiente, sin embargo el algoritmo es capaz de readaptarse y seguir cumpliendo su función aunque no sea de forma perfecta.

V.3. Arquitecturas y entrenamiento del clasificador

Como se mencionó en el capítulo [IV](#) la selección de la arquitectura para el clasificador consistió en una investigación exhaustiva para lograr la implementación de Transfer Learning para ahorrar tiempo y recursos en el diseño de la arquitectura de la red convolucional, para esto se eligieron 3 arquitecturas, ResNet50 con pesos pre-entrenados de ImageNet, VGG-16 con pesos en formato 'bottleneck features' basados en un clasificador de 133 razas de perros y por último la arquitectura MobileNet preentrenada sobre 1000 clases en ImageNet a través de la herramienta Teachable Machine ([página web de Teachable Machine](#)).

Se entrenaron dos versiones del clasificador, una versión con 4 clases en las 3 arquitecturas mencionadas y otra versión con 3 clases solo en la arquitectura de MobileNet ya que resultó ser la mejor de las tres arquitecturas usadas.

V.3.1. Modelo 4 clases

Este modelo está conformado por las siguientes clases:

- Agresividad
- Ansiedad
- Miedo
- Nada

El proceso de entrenamiento se llevó a cabo utilizando un total de 1067 imágenes repartidas entre las 4 categorías, 343 en la categoría de Agresividad, 212 en la categoría de Ansiedad, 160 en la categoría de Miedo y 352 en la categoría de Nada, mientras que para las pruebas de clasificación se utilizaron 52 imágenes en la categoría de Agresividad, 32 en la categoría de Ansiedad, 24 en la categoría de Miedo y 53 en la categoría de Nada.

Como se puede observar, las categorías de Ansiedad y Miedo son las más desbalanceadas debido a las características o posturas visuales tan específicas a considerar para clasificar dichos comportamientos en estas categorías.

Durante el entrenamiento de los modelos, el clasificador basado en ResNet50 fue el que obtuvo un peor valor de precisión o accuracy, con un valor de 0.3524 en el mejor de los casos y un 0.2948 en el peor, mientras que el valor de validation accuracy se mantenía con un valor promedio de 0.3262 constante con ligeras variaciones a 0.3302 como máximo ó 0.3209 como mínimo y este clasificador fue el que tuvo mayor problema para clasificar las imágenes de Ansiedad y Miedo, clasificando todas las imágenes de prueba pertenecientes a estas categorías en la categoría de Nada.

El modelo basado en VGG-16 obtuvo resultados similares al clasificador basado en ResNet50 durante el entrenamiento, con valores ligeramente mejores como un 0.3920 en accuracy en el mejor de los casos y un 0.3056 en el peor, por otro lado el validation accuracy obtuvo un valor promedio de 0.3425 con ligeros cambios a 0.3683 como máximo o 0.3051 como mínimo. Respecto a la clasificación, el desempeño no fue tan bueno como se esperaba, ya que presentó dificultades para clasificar las categorías de Ansiedad y Miedo, si bien no clasificó erróneamente todas las muestras como el modelo basado en ResNet50, la gran mayoría cayeron en las categorías de Agresividad o Nada, clasificando correctamente solo el 5.3% de la categoría Ansiedad y el 3.1% de la categoría Miedo.

Finalmente, el último modelo fue entrenado sobre la arquitectura MobileNet en la herramienta Teachable Machine; este modelo fue el que obtuvo mejores resultados en comparación con los otros dos, ya que obtuvo un valor de accuracy de 0.9710 mientras que el valor de validation accuracy fue de 0.6824. El modelo de 4 clases fue capaz de clasificar correctamente el 86.5% de las muestras de prueba en la categoría de Agresividad, el 71.6% en la categoría de Nada, y el 59.3% en la categoría de Ansiedad, teniendo más dificultades en la categoría de Miedo, clasificando correctamente tan solo el 29.1% de las muestras pertenecientes a esta categoría.

En la tabla III se muestra de manera resumida los resultados de las métricas utilizadas.

Tabla III. Resultados por modelo del entrenamiento de clasificadores de 4 clases

Métrica	ResNet50	VGG-16	MobileNet
Precisión	0.3524	0.3920	0.9710
Precisión de validación	0.3262	0.3425	0.6824
Precisión de prueba	0.3002	0.3243	0.6917
Macro valor F1	0.1241	0.1584	0.6046

De los tres modelos el que obtuvo mejores resultados fue el modelo entrenado con la arquitectura MobileNet, obteniendo un valor de precisión durante el entrenamiento de 0.971 comparado con un valor de 0.3524 en ResNet50 y 0.392 en VGG-16 usando la misma base de datos perteneciente al proyecto, además de que el modelo entrenado en MobileNet presentaba menos problemas en la matriz de confusión al momento de clasificar las imágenes de prueba en las cuatro categorías.

V.3.2. Modelo de 3 clases

Al ver que los clasificadores anteriores presentaron algunas dificultades para clasificar las muestras pertenecientes a las clases de Ansiedad y Miedo se optó por

probar unir dichas clases y volverla una sola, teniendo así un modelo con las siguientes clases:

- Agresividad
- Ansiedad/Miedo
- Nada

Uniendo las categorías de Ansiedad y Miedo la distribución de las imágenes quedó de 343 en la categoría de Agresividad, 372 en la categoría unificada de Ansiedad/Miedo y 352 en la categoría de Nada, dando como resultado un mejor balance entre las tres clases.

Para el entrenamiento de este modelo se decidió utilizar únicamente la arquitectura de MobileNet bajo la herramienta Teachable Machine, ya que fue la que arrojó los mejores resultados durante el entrenamiento de 4 clases y obtuvo el mejor desempeño al momento de la clasificación. Los valores de accuracy o precisión obtenidos durante el entrenamiento fue de 0.9921 mientras que en validation accuracy obtuvo un 0.6861.

Si bien los valores de accuracy y validation accuracy fueron muy similares al modelo de 4 clases, se observó un cambio más notable en la clasificación. Se calculó la precisión por clase tanto del modelo de 4 clases como del modelo de 3 clases, obteniendo los siguientes resultados:

Tabla IV. Comparación de precisión por clase en los modelos de 4 clases y de 3 clases

Modelo	Agresividad	Ansiedad	Miedo	Nada
Modelo 4 clases	0.87	0.59	0.29	0.72
Modelo 3 clases	0.71	0.64		0.64

Durante la clasificación de las muestras de prueba, el modelo de 3 clases clasificó correctamente el 71.1 % de las muestras de Agresividad, el 64.2 % de las muestras de la

categoría unificada Ansiedad/Miedo y el 64.1 % en la categoría Nada. En comparación con el modelo de 4 clases, el porcentaje de aciertos en la clasificación de las categorías de Agresividad y Nada disminuyó de 86.5 % y 71.6 % a 71.1 % y 64.1 % respectivamente en el modelo de 3 clases, por otro lado en la categoría de Ansiedad y Miedo se obtuvieron valores más balanceados respecto a las clases restantes, ya que al estar separadas dichas clases se obtuvo un porcentaje de aciertos de clasificación del 59.3 % y 29.1 % respectivamente, y al ser unificadas en una única clase se obtuvo un porcentaje de aciertos del 64.2 %.

El resumen de porcentajes se pueden ver en las tablas [V](#) y [VI](#). También se puede observar la matriz de confusión de 4 clases y de 3 clases en la figuras [18](#) y [19](#) respectivamente.

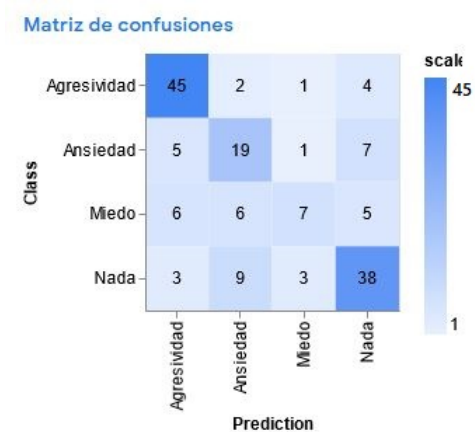


Figura 18. Matriz de confusiones modelo de 4 clases

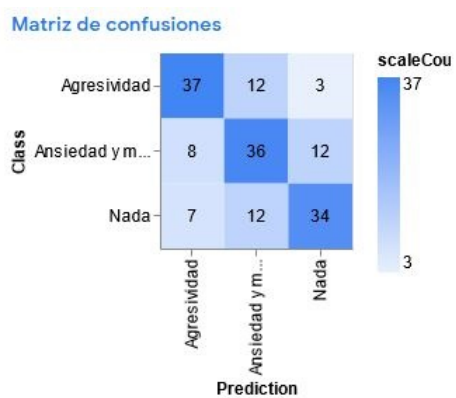


Figura 19. Matriz de confusiones modelo de 3 clases

Tabla V. Porcentaje de aciertos durante la prueba de clasificación: modelo 4 clases

	Agresividad	Ansiedad	Miedo	Nada
Porcentaje	86.5%	59.3%	29.1%	71.6%

Tabla VI. Porcentaje de aciertos durante la prueba de clasificación: modelo 3 clases

	Agresividad	Ansiedad/Miedo	Nada
Porcentaje	71.1%	64.2%	64.1%

Capítulo VI

Diseño del prototipo

En este capítulo se presenta el diseño de un prototipo de aplicación práctica a manera de aplicación web que es la manera en la que se presentan los resultados del análisis al usuario común.

La ventana principal consiste una interfaz sencilla con una imagen de bienvenida, un texto con las instrucciones a seguir durante la ejecución, una imagen de bienvenida, y un botón con el texto "Analizar muestras" como se observa en la figura [20](#).



Figura 20. Ventana principal de la aplicación

Una vez que el usuario presiona el botón de "Analizar muestras" se pide acceso a la cámara web del usuario, a lo cual el usuario debe permitir el acceso para proceder con el análisis, esto se puede apreciar mejor en la figura [21](#).



Figura 21. Ventana emergente para solicitar acceso a la cámara web

Posteriormente el sistema abre un pequeño recuadro donde se muestra lo que está captando la cámara web, llegado a este punto el usuario puede mostrar imágenes de perros que estén presentando alguna de las actitudes identificables por el modelo delante de la cámara web.

Además, justo debajo del recuadro, el sistema muestra los valores de la probabilidad de que la imagen mostrada pertenezca a una categoría u otra según lo que pueda captar. En las figuras [22](#) y [23](#) se puede observar lo descrito anteriormente.



Figura 22. Cámara web abierta y lista para el análisis

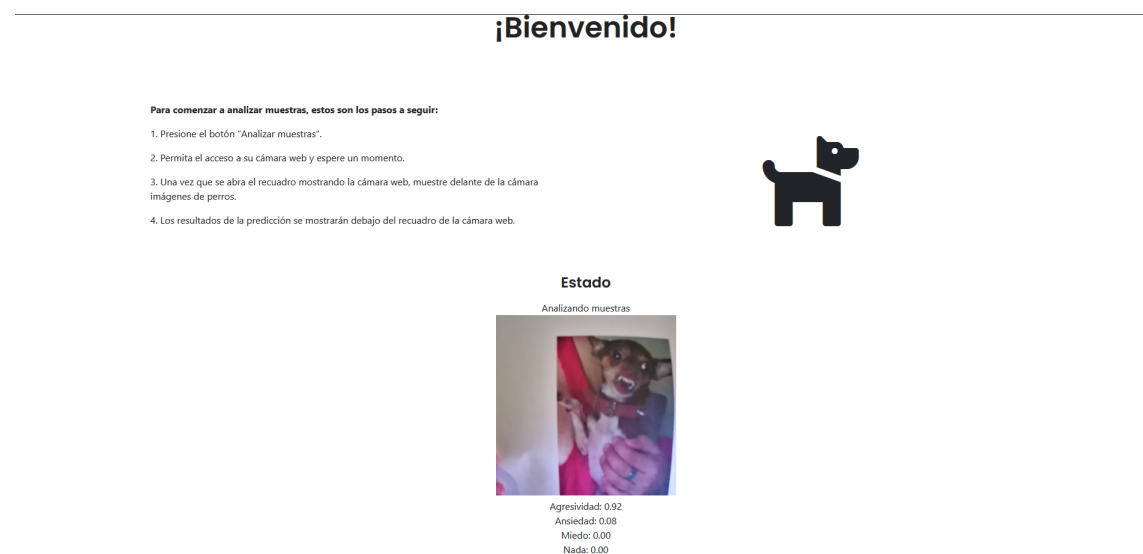


Figura 23. Análisis en tiempo real de una imagen mostrada frente a la cámara

El usuario puede seguir mostrando imágenes con posturas diferentes y el sistema seguirá realizando las predicciones pertinentes, mostrando los cambios en tiempo real. Esto se ilustra mejor en las figuras [24](#), [25](#) y [26](#).

Para comenzar a analizar muestras, estos son los pasos a seguir:

1. Presione el botón "Analizar muestras".
2. Permita el acceso a su cámara web y espere un momento.
3. Una vez que se abra el recuadro mostrando la cámara web, muestre delante de la cámara imágenes de perros.
4. Los resultados de la predicción se mostrarán debajo del recuadro de la cámara web.



Figura 24. Análisis de una imagen de la categoría Miedo

Para comenzar a analizar muestras, estos son los pasos a seguir:

1. Presione el botón "Analizar muestras".
2. Permita el acceso a su cámara web y espere un momento.
3. Una vez que se abra el recuadro mostrando la cámara web, muestre delante de la cámara imágenes de perros.
4. Los resultados de la predicción se mostrarán debajo del recuadro de la cámara web.



Figura 25. Análisis de una imagen de la categoría Nada

Para comenzar a analizar muestras, estos son los pasos a seguir:

1. Presione el botón "Analizar muestras".
2. Permita el acceso a su cámara web y espere un momento.
3. Una vez que se abra el recuadro mostrando la cámara web, muestre delante de la cámara imágenes de perros.
4. Los resultados de la predicción se mostrarán debajo del recuadro de la cámara web.



Estado

Analizando muestras



Agresividad: 0.00
Ansiedad: 1.00
Miedo: 0.00
Nada: 0.00

Figura 26. Análisis de una imagen de la categoría Ansiedad

Capítulo VII

Conclusiones

En este trabajo de tesis se desarrolló un método para la clasificación de comportamientos de perros haciendo uso de técnicas de aprendizaje máquina como fue el transfer learning, las redes neuronales convolutivas y visión artificial. Haciendo uso de estas técnicas se logró diseñar un modelo que fuera capaz de evaluar las características de un perro dadas una serie de imágenes y que a partir de dicha evaluación se extraiga información que pueda ser de utilidad durante el proceso de selección y entrenamiento de perros de rescate bajo los criterios de los entrenadores expertos.

Los clasificadores creados haciendo uso de redes neuronales convolutivas desde cero requieren una gran cantidad de datos y un poder de cómputo muy elevado para llevar a cabo el entrenamiento del modelo de manera rápida y eficiente, además de que gran parte del éxito en la extracción de características de bajo y alto nivel dependerá de la arquitectura y la configuración de la red neuronal, por lo que una opción más viable es utilizar transfer learning para entrenar un clasificador orientado a un nuevo problema específico a partir de un modelo base previamente entrenado que posea características similares al nuevo problema. Así la cantidad de datos requeridos y el poder computacional necesario se reducen de manera drástica, permitiendo crear clasificadores en equipos de cómputo convencionales y además utilizar arquitecturas de redes neuronales eficientes ya probadas en problemas de clasificación y extracción de características.

Sin embargo como se pudo observar a lo largo del trabajo, el uso de transfer learning

no siempre se obtienen los mejores resultados, esto se puede deber a muchos factores, como por ejemplo que las categorías a clasificar compartan características similares entre si o que las características a extraer sean demasiado puntuales y difíciles de detectar. También puede suceder que la arquitectura pre-entrenada y los pesos pre-calculados no siempre sean óptimos para resolver un nuevo problema aún si se comparten similitudes entre las categorías originales pre-entrenadas y las nuevas categorías que corresponden al problema específico, esto se pudo observar cuando se adaptaron al problema de este proyecto las redes neuronales pre-entrenadas ResNet50 y Vgg-16 ya que a pesar de compartir ciertas similitudes entre el problema particular de este proyecto eso no bastó para que los resultados de clasificación fuesen óptimos, por lo que si se cuenta con el poder de cómputo suficiente y una gran cantidad de datos lo más recomendable es entrenar los modelos desde cero, cargando una arquitectura robusta pero utilizando pesos aleatorios que se vayan ajustando al problema particular.

A pesar de dichas dificultades fue posible configurar y adaptar una arquitectura pre-entrenada para que resolviera de manera cuando menos decente el problema particular que se abordó a lo largo de este trabajo que fue la provista por Teachable Machine, permitiendo así diseñar una solución que combine el estudio del comportamiento canino con un modelo computacional capaz de identificar y analizar dicho comportamiento. Con esto se logra aportar una pizca de conocimiento al área de las ciencias computacionales, ya que tomando en cuenta la literatura analizada en el estado del arte la mayoría de los trabajos se centran solo en clasificar posturas básicas de los perros o en estudiar un único comportamiento sin hacer un análisis más profundo.

Así pues este trabajo aporta una herramienta que puede proporcionar información básica sobre el comportamiento de un perro en base a 4 comportamientos específicos que fueron Agresividad, Ansiedad/Nerviosismo, Miedo y Nada, ya que estos 4

comportamientos tienen una gran relevancia al momento de evaluar a perros que son candidatos para convertirse en perros de rescate, por lo que ahora un entrenador de perros de rescate puede sumar la información extraíble mediante esta herramienta y tener más datos sobre si un perro es apto o no para desempeñar labores de búsqueda y rescate.

Dicho esto es importante señalar que aún quedan muchos aspectos que mejorar, tanto a nivel metodológico como tecnológico, ya que aún es posible mejorar el modelo para que proporcione resultados más fiables y precisos que puedan otorgar información con mayor valor para la toma de decisiones en la selección y entrenamiento de perros de rescate, además de que todavía se puede mejorar el método de clasificación y extracción de características ya que algunos comportamientos son más difíciles de detectar que otros debido a lo específico de los movimientos o de las acciones que se usan para determinar dichos comportamientos.

En conclusión, el estado actual del método propuesto proporciona una herramienta capaz de analizar y clasificar el comportamiento del perro doméstico limitándose a 4 comportamientos básicos que son agresividad, ansiedad/nerviosismo, miedo y nada, esto se logró haciendo uso de técnicas de aprendizaje máquina, visión artificial y transfer learning. Con ello la herramienta proporciona información acerca de qué tan probable es que un perro presente uno de los comportamientos descritos, y con esto se proveen datos que pueden resultar de utilidad para que los entrenadores de perros de rescate tomen decisiones con más elementos disponibles acerca de si un perro es apto o no para labores de rescate. Además se deja en claro que existe la posibilidad de mejora a nivel metodológico y tecnológico, con lo que este modelo puede servir como base para futuras investigaciones que tengan como finalidad la clasificación del comportamiento canino mediante visión artificial o técnicas de aprendizaje máquina y que cuenten con

más datos o mejor poder computacional.

Referencias bibliográficas

- Al-Amri, S. S., Kalyankar, N., y Khamitkar, S. (2010). Image segmentation by using edge detection. *International journal on computer science and engineering*, **2**(3): 804–807.
- Amir, S., Zamansky, A., y van der Linden, D. (2017). K9-blyzer: Towards video-based automatic analysis of canine behavior. En *Proceedings of the Fourth International Conference on Animal-Computer Interaction*, pp. 1–5.
- Ayodele, T. O. (2010). Types of machine learning algorithms, new advances in machine learning, yagang zhang (ed.), intech, 2010, doi: 10.5772/9385. Available from: <http://www.intechopen.com/books/new-advances-in-machinelearning/types-of-machinelearning-algorithms>.
- Barros, P., Magg, S., Weber, C., y Wermter, S. (2014). A multichannel convolutional neural network for hand posture recognition. En *International Conference on Artificial Neural Networks*, pp. 403–410. Springer.
- Bishop, C. M. *et al.* (1995). *Neural networks for pattern recognition*. Oxford university press.
- Brugarolas, R., Roberts, D., Sherman, B., y Bozkurt, A. (2013). Machine learning based posture estimation for a wireless canine machine interface. En *2013 IEEE Topical Conference on Biomedical Wireless Technologies, Networks, and Sensing Systems*, pp. 10–12. IEEE.
- Brunelli, R. (2009). *Template matching techniques in computer vision: theory and practice*. John Wiley & Sons.
- Carney, M., Webster, B., Alvarado, I., Phillips, K., Howell, N., Griffith, J., Jongejan, J., Pitaru, A., y Chen, A. (2020). Teachable machine: Approachable web-based tool for exploring machine learning classification. En *Extended abstracts of the 2020 CHI conference on human factors in computing systems*, pp. 1–8.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., y Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. En *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. Ieee.
- Doan, H.-G., Nguyen, V.-T., Vu, H., y Tran, T.-H. (2016). A combination of user-guide scheme and kernel descriptor on rgb-d data for robust and realtime hand posture recognition. *Engineering Applications of Artificial Intelligence*, **49**: 103–113.
- Dugatkin, L. A. (2020). *Principles of animal behavior*. University of Chicago Press.

- Ehsani, K., Bagherinezhad, H., Redmon, J., Mottaghi, R., y Farhadi, A. (2018). Who let the dogs out? modeling dog behavior from visual data. En *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4051–4060.
- Elforaici, M. E. A., Chaaraoui, I., Bouachir, W., Ouakrim, Y., y Mezghani, N. (2018). Posture recognition using an rgb-d camera: exploring 3d body modeling and deep learning approaches. En *2018 IEEE life sciences conference (LSC)*, pp. 69–72. IEEE.
- Ghosh, A., Sufian, A., Sultana, F., Chakrabarti, A., y De, D. (2020). Fundamental concepts of convolutional neural network. En *Recent Trends and Advances in Artificial Intelligence and Internet of Things*, pp. 519–567. Springer.
- Goldberg, D. E. y Holland, J. H. (1988). Genetic algorithms and machine learning.
- Gonzalez, R. C., Woods, R. E., *et al.* (2002). Digital image processing.
- Handelman, B. (2012). *Canine behavior: A photo illustrated handbook*. Dogwise Publishing.
- He, K., Zhang, X., Ren, S., y Sun, J. (2016). Deep residual learning for image recognition. En *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.
- Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., y Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.
- Ibraheem, N. A., Hasan, M. M., Khan, R. Z., y Mishra, P. K. (2012). Understanding color models: a review. *ARPJ Journal of science and technology*, **2**(3): 265–275.
- Jones, A. C. y Gosling, S. D. (2005). Temperament and personality in dogs (*canis familiaris*): A review and evaluation of past research. *Applied Animal Behaviour Science*, **95**(1-2): 1–53.
- Jordan, M. I. y Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, **349**(6245): 255–260.
- Lindblad-Toh, K., Wade, C. M., Mikkelsen, T. S., Karlsson, E. K., Jaffe, D. B., Kamal, M., Clamp, M., Chang, J. L., Kulbokas, E. J., Zody, M. C., *et al.* (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature*, **438**(7069): 803–819.
- Lowe, D. G. (1999). Object recognition from local scale-invariant features. En *Proceedings of the seventh IEEE international conference on computer vision*, Vol. 2, pp. 1150–1157. Ieee.
- Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, **60**(2): 91–110.

- Man, D. y Vision, A. (1982). A computational investigation into the human representation and processing of visual information.
- Mealin, S., Domínguez, I. X., y Roberts, D. L. (2016a). Semi-supervised classification of static canine postures using the microsoft kinect. En *Proceedings of the Third International Conference on Animal-Computer Interaction*, pp. 1–4.
- Mealin, S., Howell, S., y Roberts, D. L. (2016b). Towards unsupervised canine posture classification via depth shadow detection and infrared reconstruction for improved image segmentation accuracy. En *Conference on Biomimetic and Biohybrid Systems*, pp. 155–166. Springer.
- Michie, D. (1968). “memo” functions and machine learning. *Nature*, **218**(5136): 19–22.
- Mukai, N., Zhang, Y., y Chang, Y. (2018). Pet face detection. En *2018 Nicograph International (NicoInt)*, pp. 52–57. IEEE.
- Nasirahmadi, A., Sturm, B., Edwards, S., Jeppsson, K.-H., Olsson, A.-C., Müller, S., y Hensel, O. (2019). Deep learning and machine vision approaches for posture detection of individual pigs. *Sensors*, **19**(17): 3738.
- O’Shea, K. y Nash, R. (2015). An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*.
- Osisanwo, F., Akinsola, J., Awodele, O., Hinmikaiye, J., Olakanmi, O., y Akinjobi, J. (2017). Supervised machine learning algorithms: classification and comparison. *International Journal of Computer Trends and Technology (IJCTT)*, **48**(3): 128–138.
- Papari, G. y Petkov, N. (2011). Edge and line oriented contour detection: State of the art. *Image and Vision Computing*, **29**(2-3): 79–103.
- Pozos-Estrada, A. A., GÁMartÁnez, R., y Hong, H. (2014). DesagregaciÃdel peligro sÃsmico para algunos sitios seleccionados de MÃ. *IngenierÃa sÃsmica*, pp. 31 – 53.
- Raju, P. D. R. y Neelima, G. (2012). Image segmentation by using histogram thresholding. *International Journal of Computer Science Engineering and Technology*, **2**(1): 776–779.
- Redmon, J. (2013–2016). Darknet: Open source neural networks in c. <http://pjreddie.com/darknet/>.
- Redmon, J. y Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.
- Ribeiro, C., Ferworn, A., Denko, M., y Tran, J. (2009). Canine pose estimation: A computing for public safety solution. En *2009 Canadian Conference on Computer and Robot Vision*, pp. 37–44. IEEE.

- Robinson, C., Mancini, C., van der Linden, J., Guest, C., y Swanson, L. (2015). Exploring assistive technology for assistance dog owners in emergency situations. En *Proceedings of the 8th ACM International Conference on Pervasive Technologies Related to Assistive Environments*, pp. 1–2.
- Simonyan, K. y Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Sinha, D. y El-Sharkawy, M. (2019). Thin mobilenet: An enhanced mobilenet architecture. En *2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, pp. 0280–0285. IEEE.
- Sucar, L. E. y Gómez, G. (2011). Visión computacional. *Instituto Nacional de Astrofísica, Óptica y Electrónica. México*.
- Tillett, R. D., Onyango, C. M., y Marchant, J. (1997). Using model-based image processing to track animal movements. *Computers and electronics in agriculture*, **17**(2): 249–261.
- Torrey, L. y Shavlik, J. (2010). Transfer learning. En *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, pp. 242–264. IGI global.
- Uher, J. y Asendorpf, J. B. (2008). Personality assessment in the great apes: Comparing ecologically valid behavior measures, behavior ratings, and adjective ratings. *Journal of Research in Personality*, **42**(4): 821–838.
- Wayne, R. K. (1993). Molecular evolution of the dog family. *Trends in genetics*, **9**(6): 218–224.
- Wilson, D. E. y Reeder, D. M. (2005). *Mammal species of the world: a taxonomic and geographic reference*, Vol. 1. JHU Press.
- Zoph, B., Cubuk, E. D., Ghiasi, G., Lin, T.-Y., Shlens, J., y Le, Q. V. (2020). Learning data augmentation strategies for object detection. En *European Conference on Computer Vision*, pp. 566–583. Springer.