

UNIVERSIDAD MAYOR DE SAN ANDRÉS

FACULTAD DE CIENCIAS PURAS Y NATURALES

CARRERA DE INFORMÁTICA



Modelo de Conducción Autónoma Basado en Aprendizaje Profundo y
Algoritmos de Visión Computacional

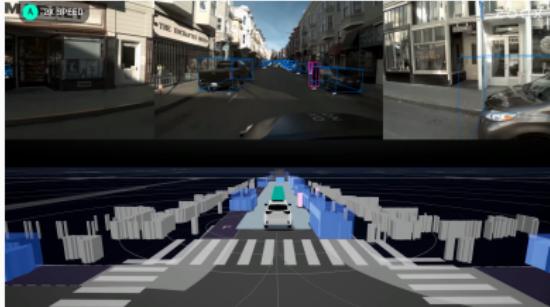
Rafael Vilca Poggian

MARCO REFERENCIAL

Introducción



Ficción



Realidad



Propuestas Alternativas

Antecedentes



Dave



Darpa Grand Challenge



Competencias de vehículos a escala

Objetivo General

“Plantear un modelo para la conducción autónoma que logre una autonomía básica en vías de doble sentido con separación física”

Objetivos Específicos

- Diseñar un componente de aumentación y preprocesamiento de datos para extraer y crear un dataset con el fin de resolver la tarea.

Objetivos Específicos

- Diseñar un componente de aumentación y preprocesamiento de datos para extraer y crear un dataset con el fin de resolver la tarea.
- Reducir la complejidad de implementación del modelo mediante el uso de solamente una cámara.

Objetivos Específicos

- Diseñar un componente de aumentación y preprocesamiento de datos para extraer y crear un dataset con el fin de resolver la tarea.
- Reducir la complejidad de implementación del modelo mediante el uso de solamente una cámara.
- Modificar y entrenar redes neuronales con una alta exactitud en las predicciones utilizando menos requisitos de cómputo.

Objetivos Específicos

- Analizar las predicciones de los modelos entrenados para comprobar si las representaciones aprendidas son invariantes a los cambios de perspectiva, iluminación y objetos en la imagen.

Objetivos Específicos

- Analizar las predicciones de los modelos entrenados para comprobar si las representaciones aprendidas son invariantes a los cambios de perspectiva, iluminación y objetos en la imagen.
- Combinar las salidas de algoritmos de visión computacional y modelos de aprendizaje profundo para mejorar la generalización de predicciones.

Objetivos Específicos

- Analizar las predicciones de los modelos entrenados para comprobar si las representaciones aprendidas son invariantes a los cambios de perspectiva, iluminación y objetos en la imagen.
- Combinar las salidas de algoritmos de visión computacional y modelos de aprendizaje profundo para mejorar la generalización de predicciones.
- Probar el rendimiento del modelo en una simulación, analizando casos de fallas y qué situaciones puede manejar correctamente.

Hipótesis

“El modelo de conducción autónoma mediante el uso de aprendizaje profundo y algoritmos de visión computacional alcanza una autonomía de nivel 2 en vías de doble sentido con separación física”

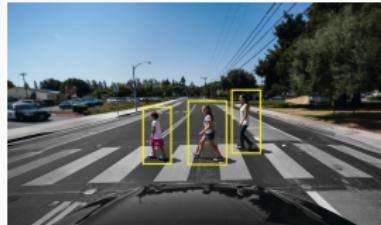
Justificación



Económica

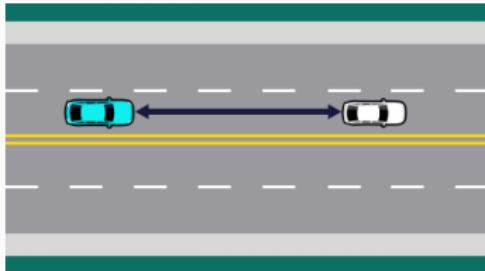


Social

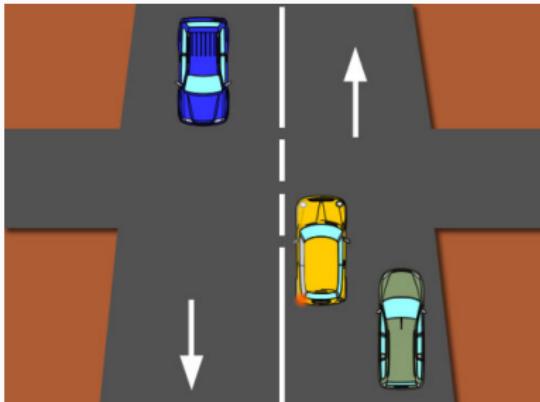
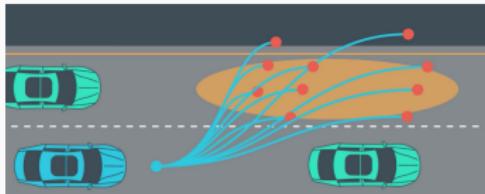
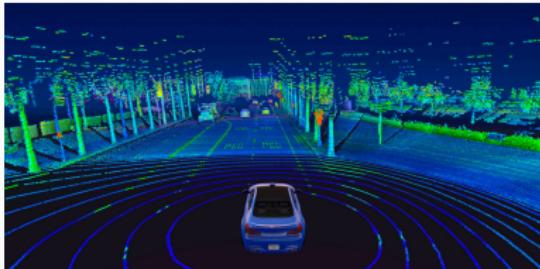


Científica

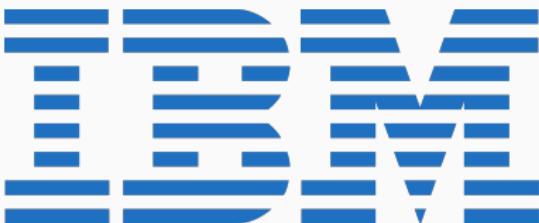
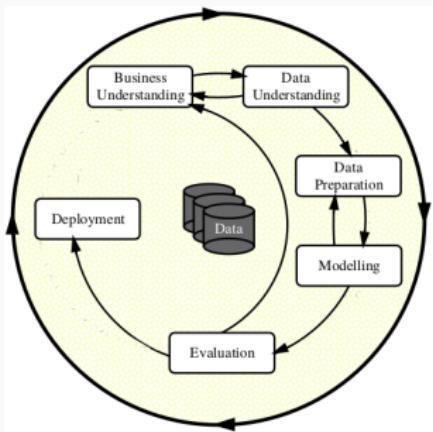
Alcances



Límites

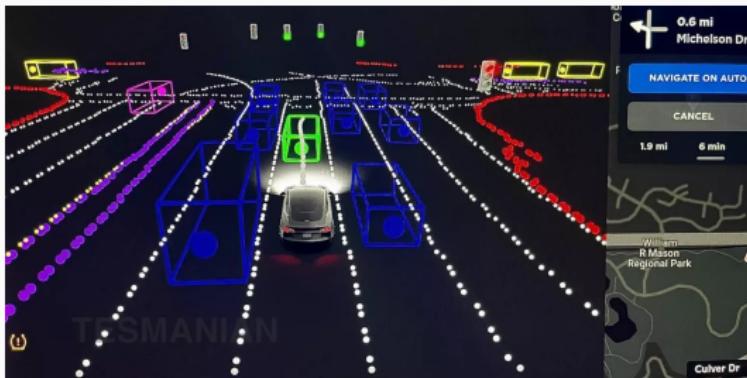


Metodología



MARCO TEÓRICO

Sistemas de Conducción Autónoma



Visión Computacional

107	98	100	105	131	115	119
18	9	11	18	46	33	42
62	55	57	61	88	71	78
97	85	76	77	131	142	120
16	13	8	14	63	61	35
59	61	59	59	104	94	66
99	88	79	83	102	127	118
21	13	6	16	34	56	43
60	56	51	57	71	86	74
122	97	94	87	110	156	128
46	15	14	16	43	67	57
82	53	53	48	74	116	91
140	127	126	115	139	162	119
60	43	42	41	66	88	44
95	76	75	68	95	121	83
152	136	135	138	133	124	110
68	56	55	62	54	42	32
102	85	82	90	86	80	74

$$\begin{array}{|c|c|c|c|c|c|c|c|} \hline
 0 & 1 & \boxed{\begin{matrix} 1 \\ \times_9 \\ 1 \\ \times_8 \\ 1 \\ \times_7 \end{matrix}} & 0 & 0 & 0 & 0 & 0 \\ \hline
 0 & 0 & \boxed{\begin{matrix} 1 \\ \times_6 \\ 1 \\ \times_5 \\ 1 \\ \times_4 \end{matrix}} & 0 & 0 & 0 & 0 & 0 \\ \hline
 0 & 0 & \boxed{\begin{matrix} 0 \\ \times_3 \\ 1 \\ \times_2 \\ 1 \\ \times_1 \end{matrix}} & 1 & 0 & 0 & 0 & 0 \\ \hline
 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 \\ \hline
 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ \hline
 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ \hline
 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline
 \end{array}$$

I

$$\begin{array}{|c|c|c|c|c|c|c|c|} \hline
 19 & 34 & 35 & 26 & 11 \\ \hline
 7 & 20 & 36 & 37 & 23 \\ \hline
 1 & 14 & 29 & 38 & 23 \\ \hline
 7 & 21 & 29 & 23 & 9 \\ \hline
 21 & 29 & 23 & 9 & 0 \\ \hline
 \end{array}$$

$I * K$

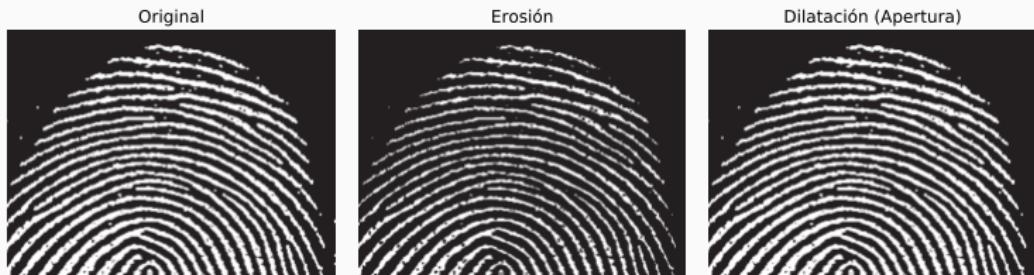
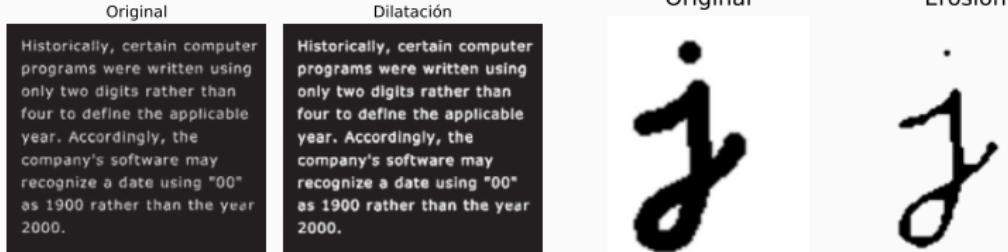
$$\begin{array}{|c|c|c|c|c|c|c|c|} \hline
 0 & 1 & 1 & \boxed{\begin{matrix} 1 \\ \times_9 \\ 0 \\ \times_8 \\ 0 \\ \times_7 \end{matrix}} & 0 & 0 & 0 \\ \hline
 0 & 0 & 1 & \boxed{\begin{matrix} 1 \\ \times_6 \\ 1 \\ \times_5 \\ 0 \\ \times_4 \end{matrix}} & 0 & 0 & 0 \\ \hline
 0 & 0 & 0 & \boxed{\begin{matrix} 1 \\ \times_3 \\ 1 \\ \times_2 \\ 1 \\ \times_1 \end{matrix}} & 0 & 0 & 0 \\ \hline
 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ \hline
 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ \hline
 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ \hline
 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline
 \end{array}$$

I

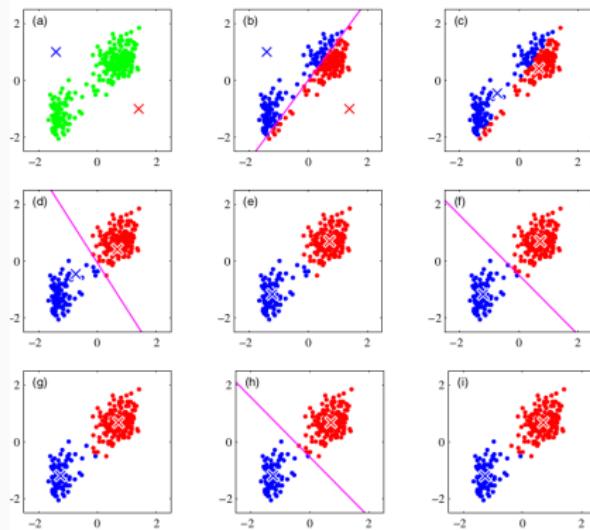
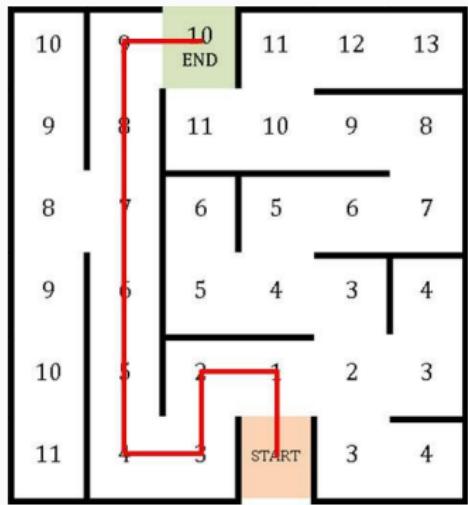
$$\begin{array}{|c|c|c|c|c|c|c|c|} \hline
 19 & 34 & 35 & 26 & 11 \\ \hline
 7 & 20 & 36 & 37 & 23 \\ \hline
 1 & 14 & 29 & 38 & 23 \\ \hline
 7 & 21 & 29 & 23 & 9 \\ \hline
 21 & 29 & 23 & 9 & 0 \\ \hline
 \end{array}$$

$I * K$

Visión Computacional



Visión Computacional



Visión Computacional



Classification



Cat

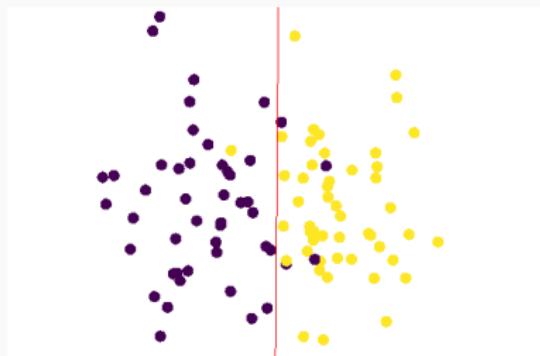
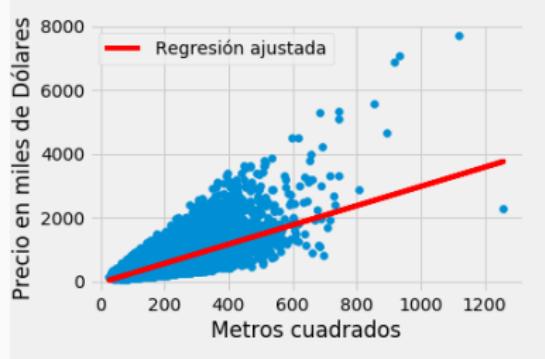
Object Detection



Semantic Segmentation

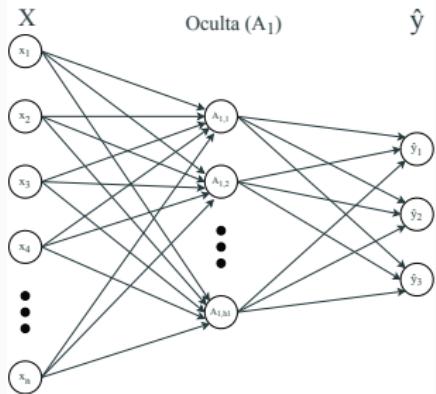


Aprendizaje Automático

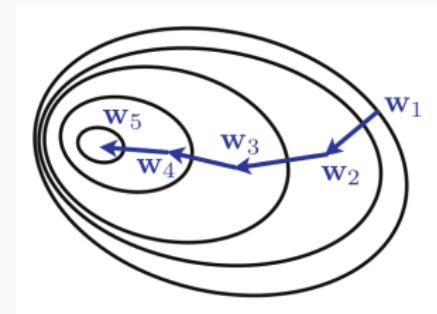
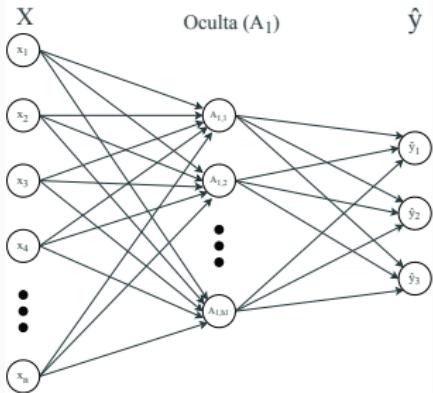


$$\text{Softmax}(Z) = \frac{e^Z}{\sum_{j=1}^C e^{Z_j}}$$

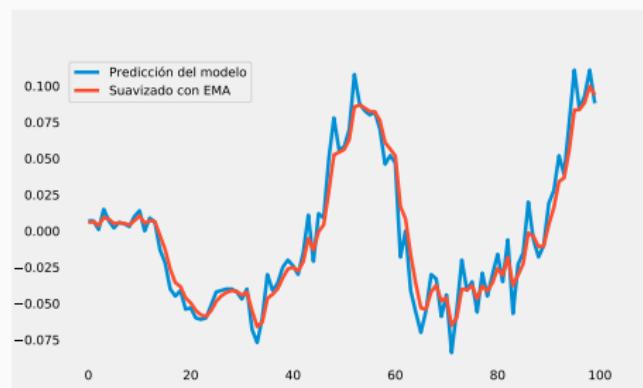
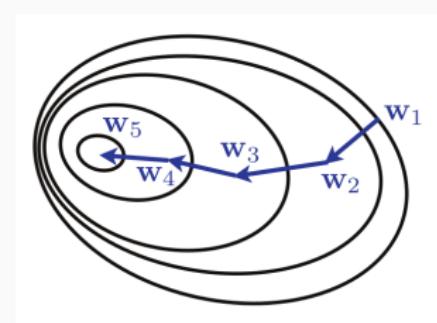
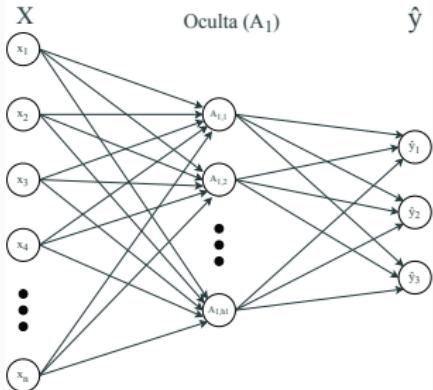
Aprendizaje Profundo



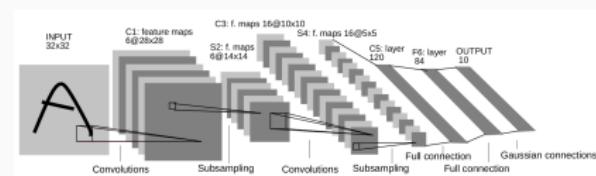
Aprendizaje Profundo



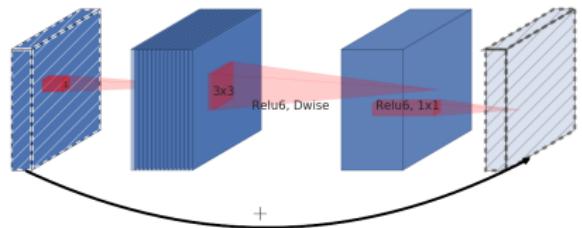
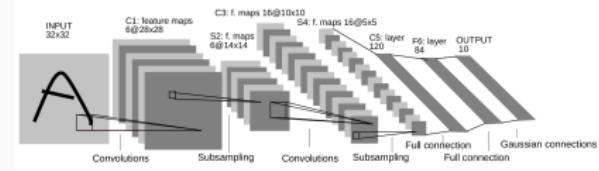
Aprendizaje Profundo



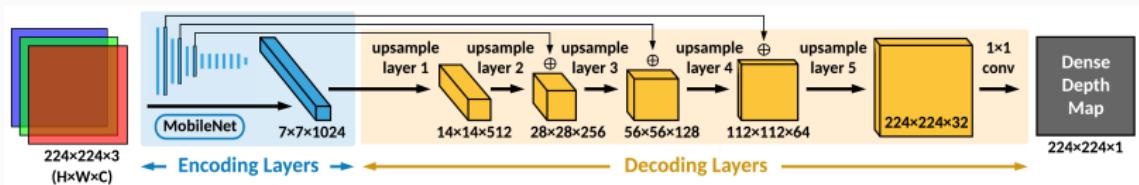
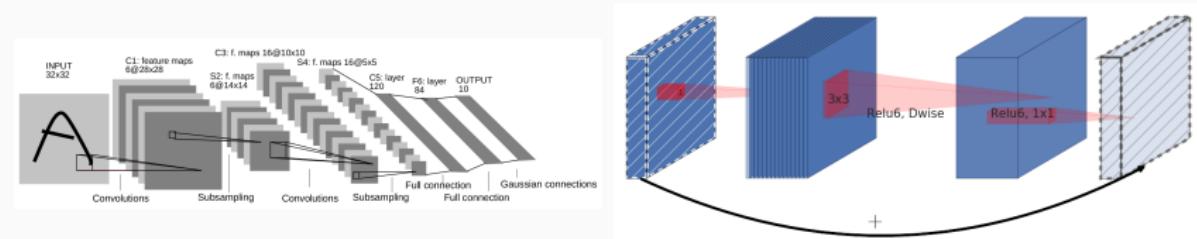
Redes Convolucionales



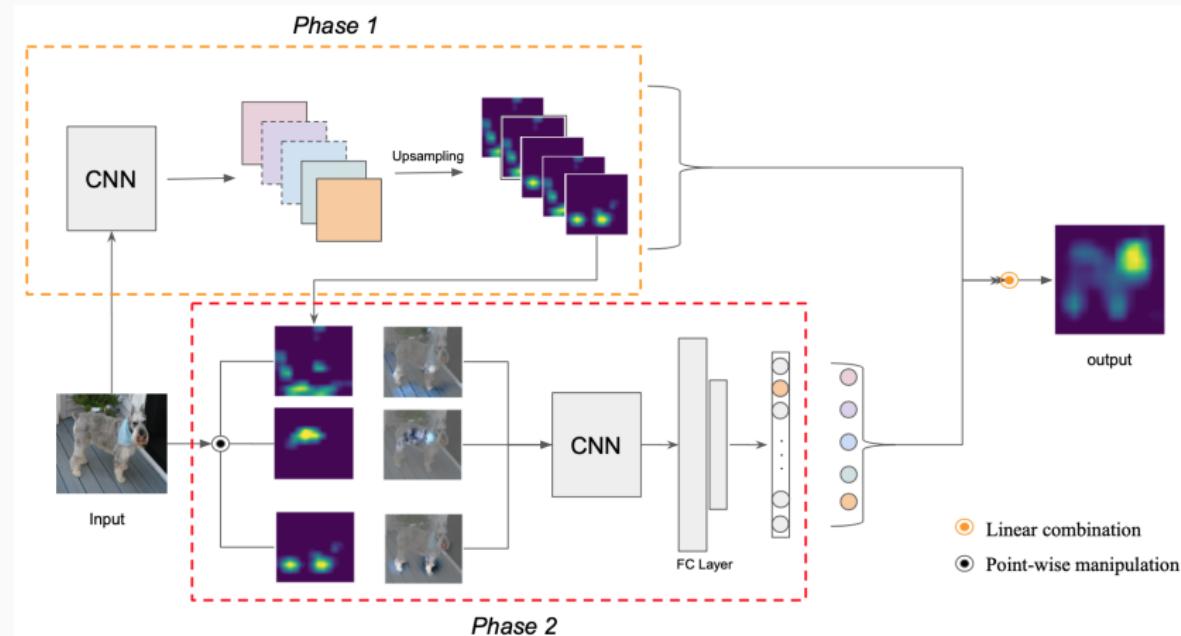
Redes Convolucionales



Redes Convolucionales



SCORECAM



Métricas de Error

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y} - y)^2$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y} - y|$$

Métricas de Error

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y} - y)^2$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y} - y|$$

$$P = \frac{T_p}{T_p + F_p}$$

$$R = \frac{T_p}{T_p + F_n}$$

$$F = 2 \frac{P \cdot R}{P + R}$$

Métricas de Error

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y} - y)^2$$

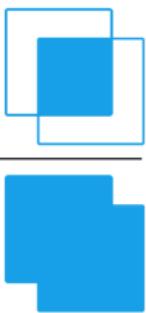
$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y} - y|$$

$$P = \frac{T_p}{T_p + F_p}$$

$$R = \frac{T_p}{T_p + F_n}$$

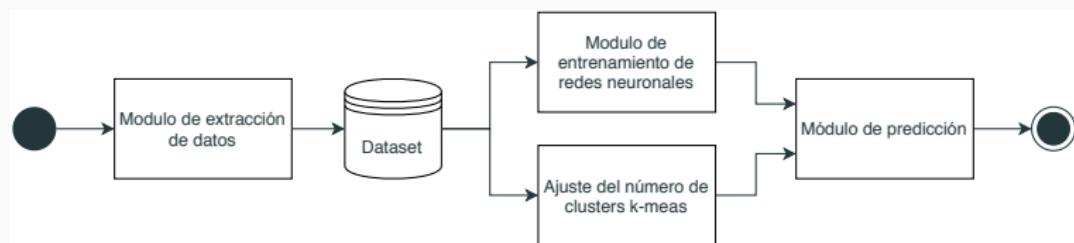
$$F = 2 \frac{P \cdot R}{P + R}$$

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

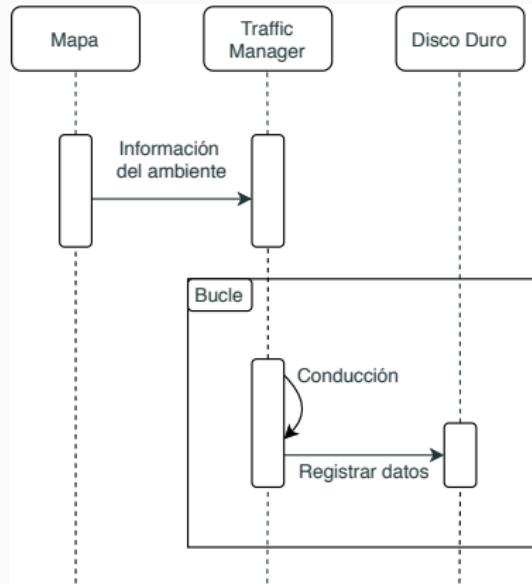


MARCO APLICATIVO

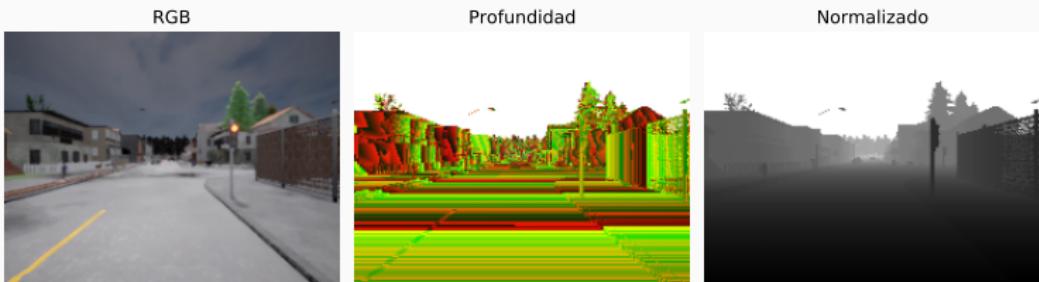
Estructura del Proyecto



Extracción de datos



Profundidad

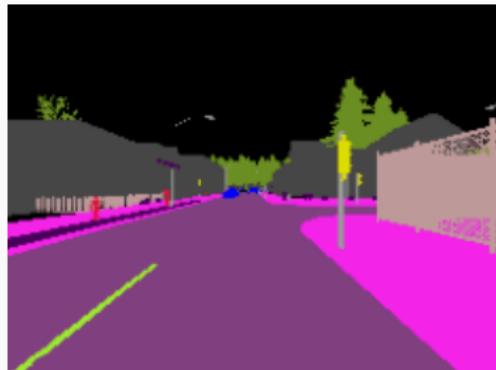


Segmentación

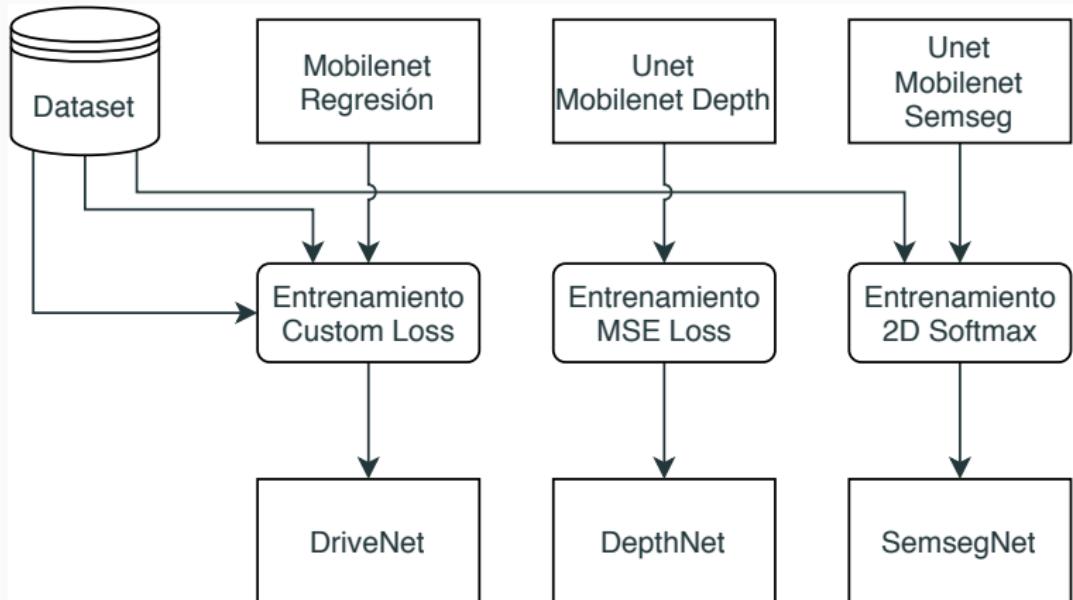
RGB



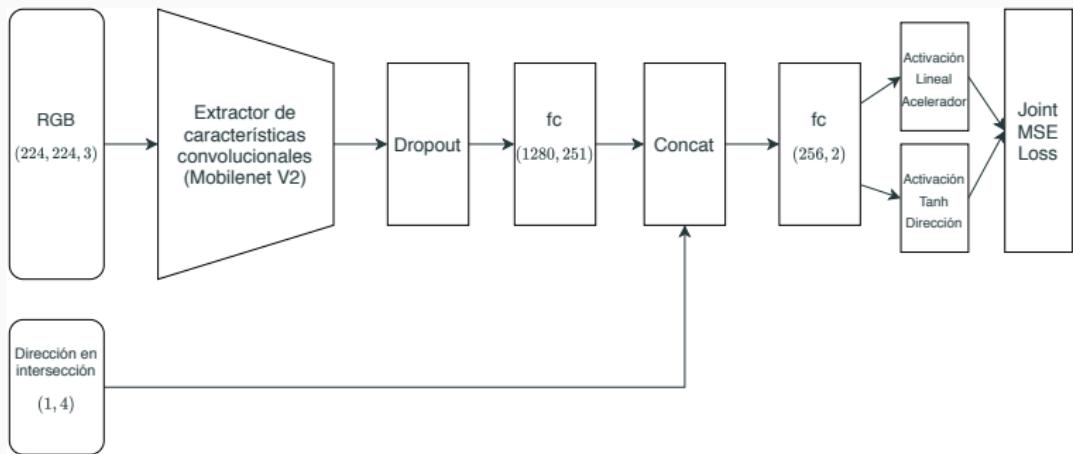
Segmentación semántica



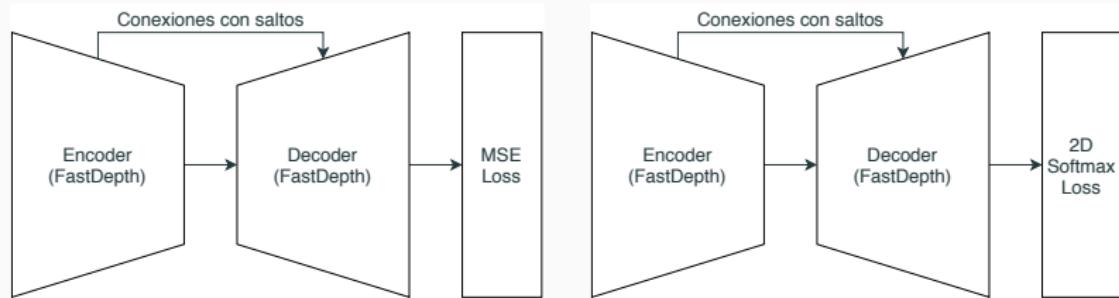
Entrenamiento de las Redes Neuronales



DriveNet



DepthNet y SemsegNet



Clasificación del Color

Verde RGB



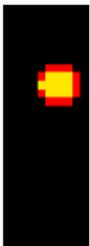
Verde Cuantizado



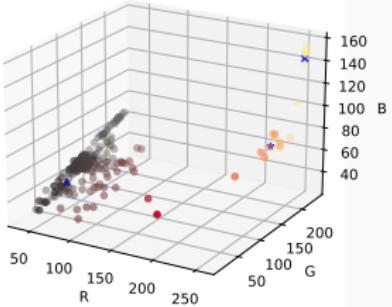
Rojo RGB



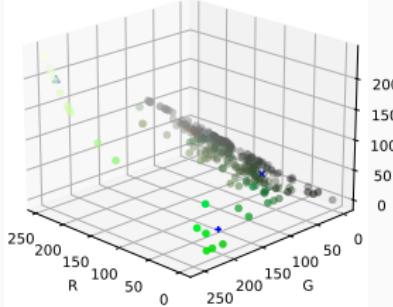
Rojo Cuantizado



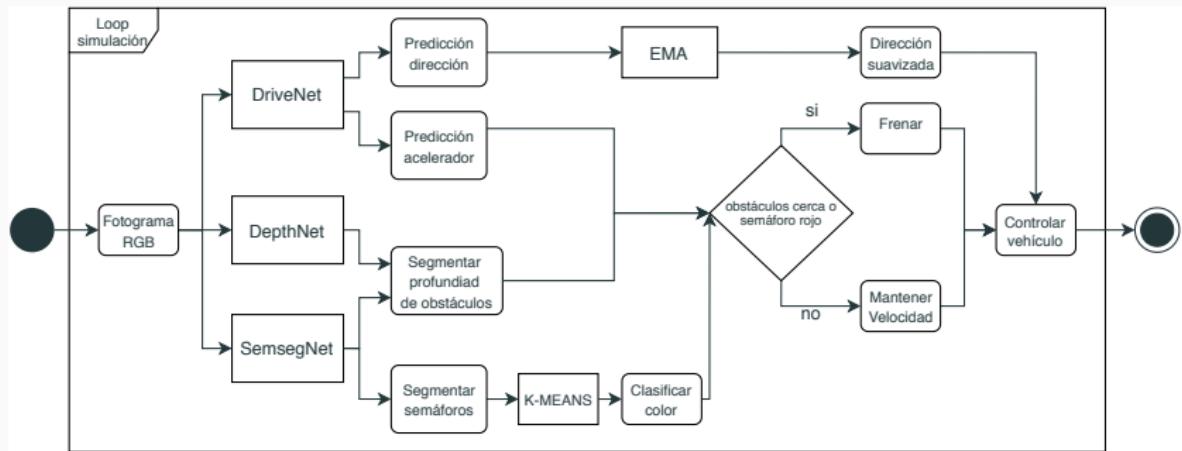
Semáforo Rojo y clusters



Semáforo Verde y clusters



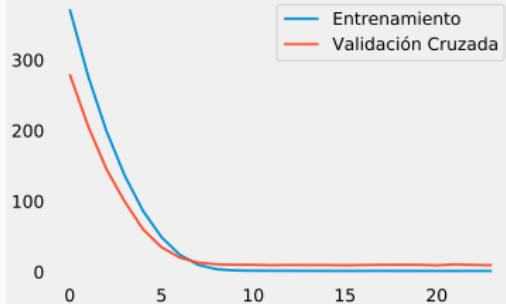
Modelo



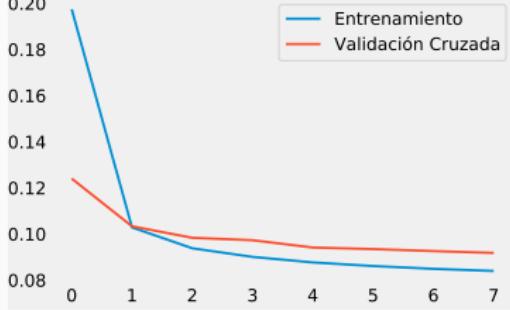
RESULTADOS Y ANÁLISIS

Curvas de Aprendizaje

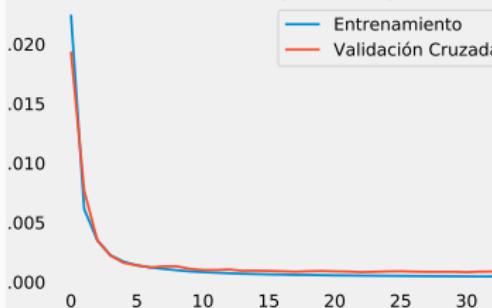
Curva de error de aprendizaje DepthNet



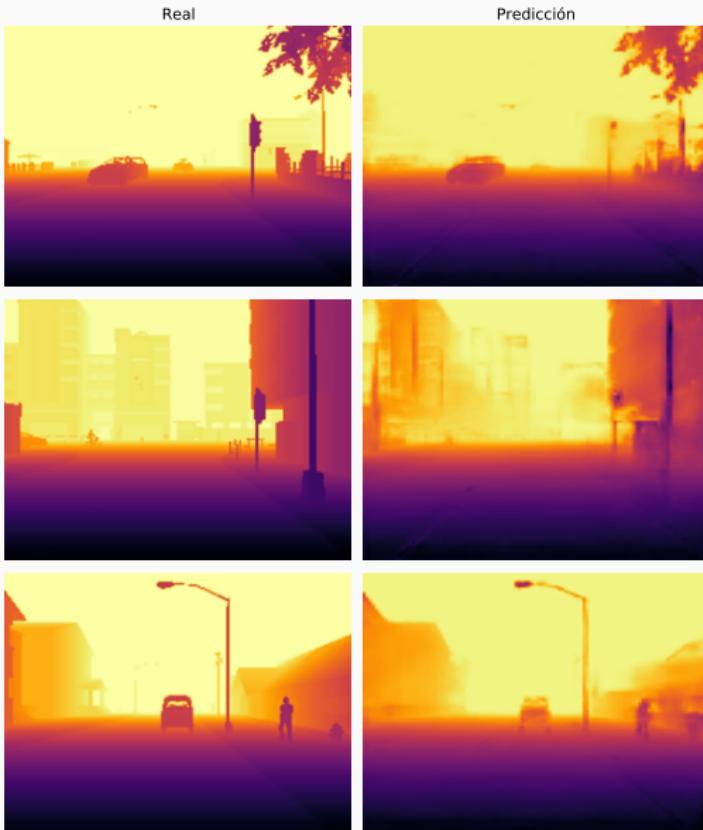
Curva de error de aprendizaje SemsegNet



Curva de error de aprendizaje DriveNet



Resultados Profundidad



Resultados Semseg

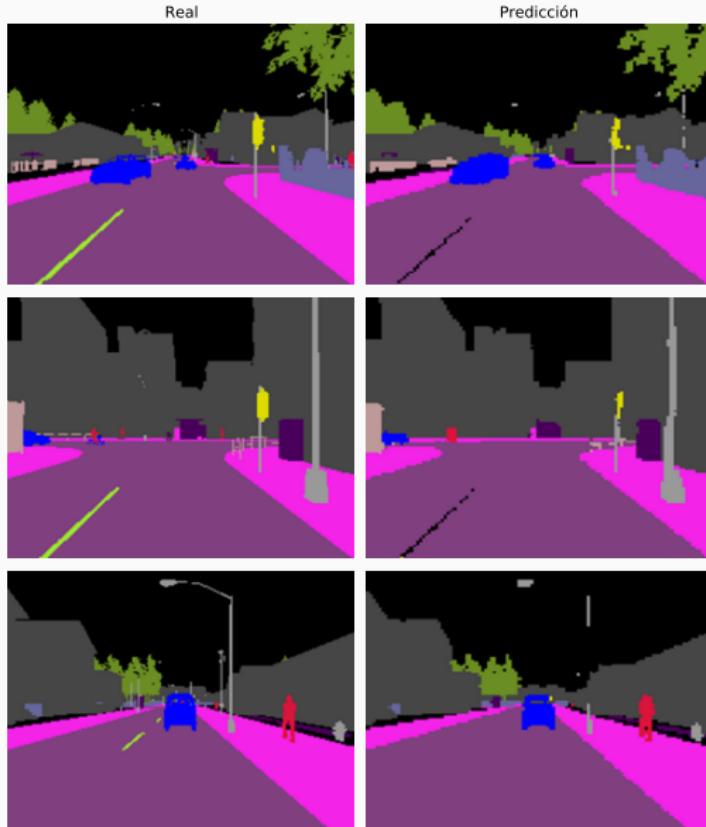
Matriz de confusión porcentual														
	Nada	Edificios	Cercas	Otro	Peatones	Postes	Carries	Caminos	Aceras	Vegetación	Vehículos	Paredes	Señales	
Nada	97.87	0.69	0.06	0.03	0.01	0.11	0.00	0.00	0.12	1.03	0.03	0.05	0.01	
Edificios	0.66	97.38	0.28	0.18	0.02	0.25	0.00	0.00	0.29	0.62	0.09	0.15	0.08	
Cercas	0.29	2.32	90.40	0.47	0.10	0.34	0.00	0.48	4.42	0.40	0.72	0.05	0.01	
Otro	0.92	8.39	1.10	80.77	0.18	0.34	0.00	0.24	5.57	0.42	0.46	1.59	0.02	
Peatones	2.00	12.47	3.64	3.30	48.25	0.50	0.00	6.50	6.75	0.14	13.76	2.70	0.00	
Postes	10.14	16.13	1.31	1.22	0.03	54.04	0.00	0.52	7.89	6.40	0.52	0.59	1.22	
Carries	37.90	0.01	0.00	0.00	0.13	0.00	0.00	61.64	0.02	0.00	0.21	0.00	0.08	
Caminos	0.01	0.00	0.03	0.01	0.01	0.01	0.00	99.13	0.58	0.00	0.22	0.00	0.00	
Aceras	0.32	0.46	0.58	0.46	0.04	0.28	0.00	1.41	96.07	0.01	0.25	0.12	0.00	
Vegetación	3.27	2.19	0.21	0.05	0.00	0.23	0.00	0.00	0.08	93.82	0.02	0.08	0.05	
Vehículos	0.19	1.33	0.51	0.19	0.24	0.03	0.00	3.72	0.98	0.16	92.54	0.04	0.07	
Paredes	1.37	4.90	0.13	0.85	0.22	0.29	0.00	0.09	1.64	0.52	0.60	89.36	0.04	
Señales	0.83	24.90	1.02	0.88	0.00	3.22	0.00	0.00	0.00	4.14	0.02	0.03	64.96	

Resultados Semseg

Clase	Precisión	Exhaustividad	Valor-F
Peatones	0.7141	0.4825	0.5759
Postes	0.7398	0.5404	0.6246
Vehículos	0.9402	0.9254	0.9327
Señales	0.7802	0.6496	0.7089

Clase	IoU
Peatones	45.72%
Postes	31.93%
Vehículos	72.46%
Señales	37.25%

Resultados Semseg

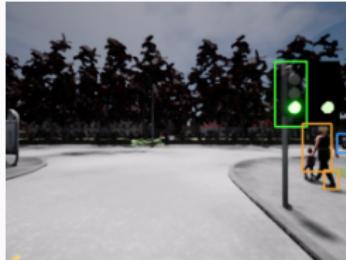


Semáforos

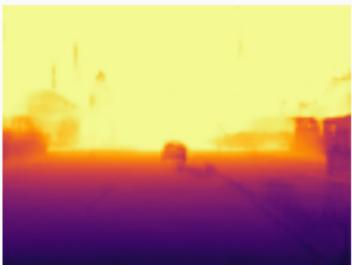
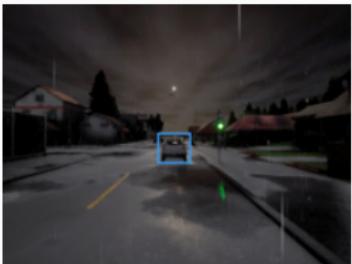
Rojos



Verdes



ROI



Zonas de Interés CNN

a) Poca iluminación



b) Sol frontal



c) Intersección



Pruebas en el Simulador

a) [real: -0.221, pred: -0.58]



b) [real: -0.283, pred: -0.8]



c) [real: -0.297, pred: -0.8]



d) [real: -0.007, pred: -0.004]



e) [real: -0.028, pred: -0.038]



f) [real: 0.002, pred: 0.016]



g) [real: -0.085, pred: 0.0]



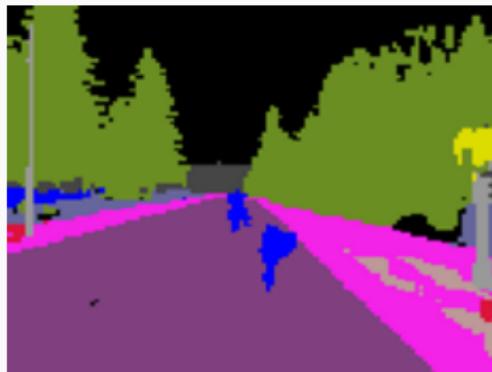
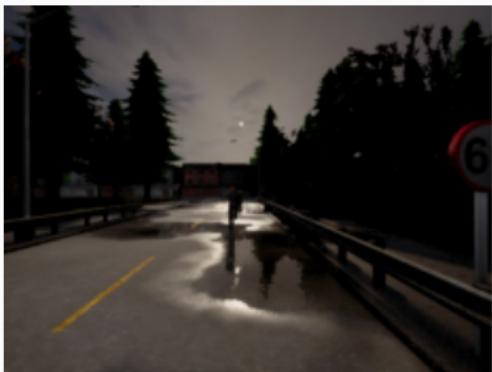
h) [real: 0.615, pred: 0.56]



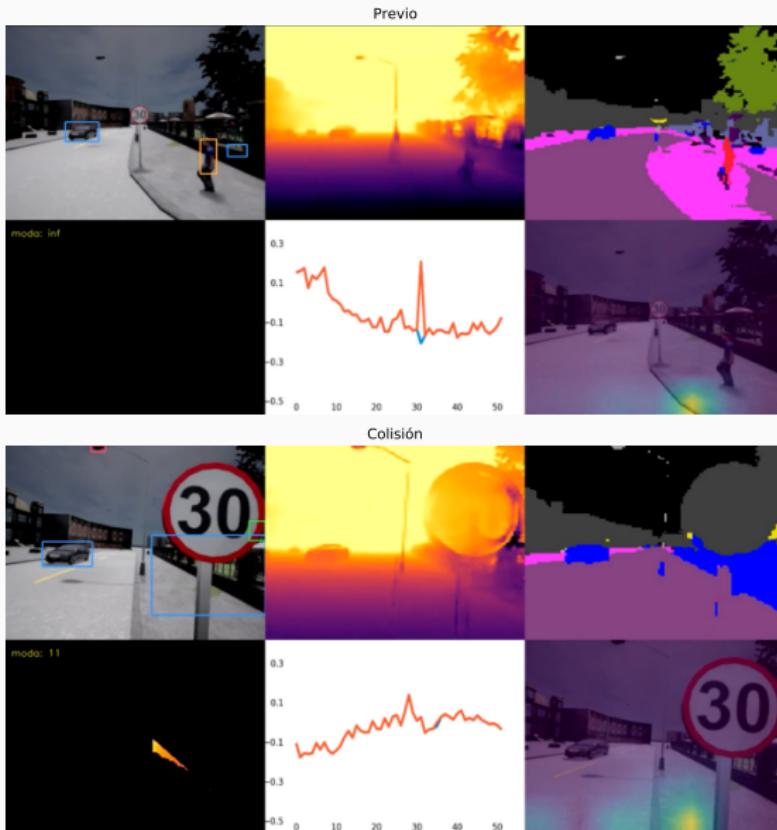
i) [real: 0.313, pred: 0.8]



Fallos



Fallos



Hipótesis

- El porcentaje de exactitud de clasificación para la detección de obstáculos es de un 71.41%

Hipótesis

- El porcentaje de exactitud de clasificación para la detección de obstáculos es de un 71.41%
- El error del control longitudinal es de ± 0.067 , considerando valores negativos como frenado y positivos como aceleración.

Hipótesis

- El porcentaje de exactitud de clasificación para la detección de obstáculos es de un 71.41%
- El error del control longitudinal es de ± 0.067 , considerando valores negativos como frenado y positivos como aceleración.
- El error de la dirección es de ± 0.069 .

Hipótesis

- El porcentaje de exactitud de clasificación para la detección de obstáculos es de un 71.41%
- El error del control longitudinal es de ± 0.067 , considerando valores negativos como frenado y positivos como aceleración.
- El error de la dirección es de ± 0.069 .
- De las 17 pruebas realizadas un 82% son satisfactorias (3 pruebas con fallos).

Hipótesis

Por lo tanto con un 93.2% de confianza se puede afirmar que se obtiene una conducción autónoma de nivel 2 en el 82% de las situaciones (17 pruebas).

CONCLUSIONES Y RECOMENDACIONES

Conclusiones

- Se diseñó un componente para la extracción, etiquetado y aumentación de datos en base a simulaciones.

Conclusiones

- Se diseñó un componente para la extracción, etiquetado y aumentación de datos en base a simulaciones.
- Se redujo la complejidad de la implementación aplicando redes neuronales y algoritmos de visión artificial sobre imágenes de una cámara RGB.

Conclusiones

- Se diseñó un componente para la extracción, etiquetado y aumentación de datos en base a simulaciones.
- Se redujo la complejidad de la implementación aplicando redes neuronales y algoritmos de visión artificial sobre imágenes de una cámara RGB.
- Se modificaron y entrenaron las redes MobileNet y FastDepth para inferir la aceleración, dirección, profundidad y segmentación semántica, con menores requisitos computacionales.

Conclusiones

- Se analizaron las predicciones en simulaciones con distintos climas, perspectivas y posiciones de obstáculos. Las predicciones son invariantes a estos cambios.

Conclusiones

- Se analizaron las predicciones en simulaciones con distintos climas, perspectivas y posiciones de obstáculos. Las predicciones son invariantes a estos cambios.
- Se combinaron las salidas de las redes neuronales con las operaciones morfológicas, eliminando residuos, además de la cuantización del color, para lograr el control dada la imagen.

Conclusiones

- Se analizaron las predicciones en simulaciones con distintos climas, perspectivas y posiciones de obstáculos. Las predicciones son invariantes a estos cambios.
- Se combinaron las salidas de las redes neuronales con las operaciones morfológicas, eliminando residuos, además de la cuantización del color, para lograr el control dada la imagen.
- Se comprobó mediante simulaciones que el modelo permite controlar la dirección y frenar ante obstáculos, cometiendo errores en situaciones puntuales.

Recomendaciones

- Diseñar una red sola red capaz de realizar la tarea de las tres redes presentadas en este trabajo.

Recomendaciones

- Diseñar una red sola red capaz de realizar la tarea de las tres redes presentadas en este trabajo.
- Extender el conjunto de datos de entrenamiento a otros mapas con distintos tipos de vías e intersecciones

Recomendaciones

- Diseñar una red sola red capaz de realizar la tarea de las tres redes presentadas en este trabajo.
- Extender el conjunto de datos de entrenamiento a otros mapas con distintos tipos de vías e intersecciones
- Incrementar la cantidad de cámaras para mejorar la percepción del ambiente y reducir puntos muertos.

Recomendaciones

- Diseñar una red sola red capaz de realizar la tarea de las tres redes presentadas en este trabajo.
- Extender el conjunto de datos de entrenamiento a otros mapas con distintos tipos de vías e intersecciones
- Incrementar la cantidad de cámaras para mejorar la percepción del ambiente y reducir puntos muertos.
- Adaptar el modelo a peatones y los objetos que se pueden encontrar en la vía pública, para ofrecer percepción artificial a personas con discapacidades visuales.