

# Tabla de Contenido

<b>Introducción</b>	<b>1</b>
<b>1. Marco teórico</b>	<b>4</b>
1.1. Definiciones . . . . .	4
1.1.1. Convolución . . . . .	4
1.1.2. Redes neuronales . . . . .	6
1.1.3. Redes convolucionales . . . . .	7
1.1.4. Algunas capas de las redes convolucionales . . . . .	7
1.1.5. Clasificación . . . . .	8
1.1.6. Arquitectura VGG . . . . .	10
1.1.7. Características y filtros . . . . .	10
1.1.8. Modelo de visualización CAM o <i>Class Activation Mapping</i> . . . . .	12
1.1.9. <i>Data augmentation</i> o Aumento de datos . . . . .	13
1.1.10. Modelos generativos . . . . .	14
1.1.11. Métricas de clasificación . . . . .	15
1.2. Problema . . . . .	16
1.3. Resumen . . . . .	17
<b>2. Propuesta</b>	<b>18</b>
2.1. Descripción de la propuesta . . . . .	18
2.1.1. ¿Por qué son difíciles de entender las redes convolucionales? . . . . .	18
2.1.2. Ayudando al modelo a aproximar la superficie objetivo . . . . .	19
2.1.3. Extraer información del modelo: Entender el modelo a partir de visualizaciones . . . . .	20
2.1.4. ¿Cómo entregar información al modelo? . . . . .	22
2.1.5. Consideraciones cruciales sobre efectividad de la propuesta . . . . .	23
2.1.6. Resultados esperados . . . . .	23
2.2. Hipótesis . . . . .	24
2.3. Preguntas de investigación . . . . .	24
2.4. Objetivos . . . . .	25
2.4.1. Objetivo General . . . . .	25
2.4.2. Objetivos Específicos . . . . .	25
2.5. Trabajo relacionado . . . . .	25
2.6. Resumen . . . . .	27
<b>3. Diseño experimental</b>	<b>28</b>

3.1.	Experimento en términos generales . . . . .	28
3.2.	Preliminares . . . . .	30
3.2.1.	<i>Dataset</i> a analizar . . . . .	30
3.2.2.	Entrenamiento inicial . . . . .	30
3.2.3.	Implementación y herramientas utilizadas . . . . .	31
3.2.4.	Límite superior del desempeño del clasificador . . . . .	33
3.2.5.	Selección de áreas irrelevantes . . . . .	33
3.3.	Métodos similares para comparación . . . . .	34
3.3.1.	Resultados . . . . .	34
3.3.2.	Análisis modelos base y métodos similares . . . . .	35
3.4.	Resumen . . . . .	36
<b>4.</b>	<b>Reajuste basado en reemplazos</b>	<b>38</b>
4.1.	Resumen de propuesta con reemplazos . . . . .	38
4.2.	Reemplazo generativo . . . . .	39
4.3.	Reemplazo recorte aleatorio . . . . .	40
4.4.	<i>Dropout</i> selectivo . . . . .	40
4.5.	Resultados reemplazos . . . . .	41
4.6.	Revisión de los supuestos . . . . .	43
4.7.	Resumen . . . . .	46
<b>5.</b>	<b>Reajuste basado en funciones de pérdida</b>	<b>47</b>
5.1.	Pérdida por activación selectiva (PAS) . . . . .	47
5.1.1.	Selección de $\lambda$ . . . . .	48
5.1.2.	Aumento de las máscaras . . . . .	49
5.1.3.	Detalle resultados PAS . . . . .	50
5.2.	Pérdida por activación selectiva ajustada (PASA) . . . . .	52
5.2.1.	Resultados pérdida PASA . . . . .	55
5.3.	Resumen . . . . .	58
<b>6.</b>	<b>Resultados para diversos <i>datasets</i></b>	<b>59</b>
6.1.	Sketches . . . . .	59
6.1.1.	Selección de máscaras . . . . .	60
6.1.2.	Resultados . . . . .	60
6.1.3.	Análisis . . . . .	61
6.2.	Experimento Símbolos . . . . .	62
6.2.1.	Resultados . . . . .	62
6.2.2.	Análisis . . . . .	63
6.3.	Experimento rayos X . . . . .	64
6.3.1.	Resultados . . . . .	65
6.3.2.	Análisis . . . . .	65
6.4.	Resumen . . . . .	67
<b>7.</b>	<b>Discusión de resultados</b>	<b>68</b>
7.1.	Exploración de los filtros de forma global . . . . .	68
7.1.1.	¿Existen características irrelevantes? . . . . .	69
7.1.2.	¿La inspección visual global encuentra las características irrelevantes? . . . . .	70

7.1.3.	¿La visualización de CAM apunta a características no relevantes?	73
7.1.4.	Efectos después de reajustar el clasificador	74
7.1.5.	Efectos al reajustar enfocado a una característica irrelevante	79
7.2.	Resumen	82
<b>Conclusión</b>		<b>82</b>
<b>Bibliografía</b>		<b>84</b>
<b>Anexo A. Derivación fórmula gradiente</b>		<b>90</b>
A.1.	Gradiente entropía cruzada respecto a <i>softmax</i>	90
A.2.	Gradiente de logits [3]	91
A.3.	Gradiente de pesos logits	92
A.4.	Gradiente de <i>global average pooling</i>	92
A.5.	Gradiente de activación filtro	92
A.6.	Gradiente de pesos <i>kernel</i> última capa	93
A.7.	Formulas resultantes	93
<b>Anexo B. Explicación general del código</b>		<b>94</b>
B.1.	Dependencias	94
B.2.	¿Cómo ejecutar código?	94
B.3.	¿Cómo conseguir los datos?	94
B.4.	Estructura del código	95
B.4.1.	<code>Classification_models</code>	95
B.4.2.	<code>Tf_recordparser</code>	95
B.4.3.	<code>Datasets</code>	95
B.4.4.	<code>ImageNet_Utils</code>	96
B.4.5.	<code>Select_tool</code>	96
B.4.6.	<code>Vis_exp</code>	96
B.4.7.	<code>Config_files</code>	96
B.4.8.	<code>Plot_utils</code>	96
B.4.9.	<code>Image_generator</code>	96
B.5.	Proceso para reproducir resultados	97