

Tabla de Contenido

1. Introducción	1
1.1. Contexto y formulación del problema	1
1.2. Objetivos	4
1.2.1. Objetivo general	4
1.2.2. Objetivos específicos	4
2. Marco teórico	5
2.1. Similitud de <i>embeddings</i>	5
2.2. Mean Average Precision (mAP)	6
2.3. Redes neuronales	7
2.3.1. Capas <i>fully-connected</i>	7
2.3.2. Capas convolucionales	8
2.3.3. Funciones de activación	9
2.3.4. Funciones de pérdida	9
2.3.5. Algoritmos de optimización	10
2.3.6. <i>Batch Normalization</i>	11
2.3.7. <i>Dropout</i>	11
2.4. ResNet	12
2.5. Semántica en la búsqueda por contenido	13
2.6. <i>Uniform Manifold Approximation and Projection</i>	14
2.7. Vectores de palabras	14
2.7.1. <i>Word2Vec</i>	14
2.7.2. <i>GloVe</i>	15
2.7.3. <i>FastText</i>	16
2.7.4. BERT (Bidirectional Encoder Representations from Transformers)	17
2.7.5. RoBERTa (Robustly Optimized BERT Pretraining Approach)	18
2.7.6. MPNet (Masked and Permuted Pre-training for Language Understanding)	19
2.8. CLIP: <i>Learning transferable visual models from natural language supervision</i>	19
3. Estudios preliminares	21
3.1. Creación de <i>datasets</i>	21
3.1.1. Dataset de Pepeganga	21
3.1.2. Dataset de IKEA	24
3.1.3. Dataset de <i>World Market</i>	27
3.1.4. Dataset de Homy	30

3.1.5. Dataset de Cartier	33
3.1.6. Dataset de UNIQLO	36
3.2. Recuperación de imágenes con ResNet-50	38
3.3. Recuperación de productos mediante texto	38
3.4. Recuperación y visualización mediante UMAP	39
4. Métodos de ajuste de modelos	40
4.1. Ajuste de <i>embeddings</i> visuales dentro del catálogo	40
4.1.1. Promedio simple con <i>moving average</i>	41
4.1.2. Promedio ponderado con similitud coseno	41
4.1.3. Promedio ponderado con <i>softmax</i>	41
4.2. <i>Query</i> adaptativa	42
4.3. Entrenamiento y recuperación utilizando el modelo CLIP	43
5. Experimentos y resultados	44
5.1. Métricas utilizadas	44
5.1.1. <i>Mean Average Precision (mAP)</i>	44
5.1.2. <i>Adjusted Rand Index</i> y <i>Adjusted Mutual Information Score</i>	44
5.2. Experimentos preliminares	44
5.2.1. Resultados y visualización de productos en 2 dimensiones	44
5.2.2. Recuperación de productos similares utilizando texto	47
5.2.2.1. Pepeganga	47
5.2.2.2. UNIQLO	48
5.2.2.3. Cartier	48
5.2.2.4. IKEA	49
5.2.2.5. World Market	49
5.2.2.6. Homy	50
5.3. Experimentos de métodos propuestos	50
5.3.1. Recuperación de imágenes utilizando ResNet-50 con ajuste de <i>embeddings</i> visuales y <i>query</i> adaptativa	50
5.3.1.1. UNIQLO	51
5.3.1.2. Cartier	53
5.3.1.3. IKEA	55
5.3.1.4. World Market	57
5.3.1.5. Homy	59
5.3.1.6. Pepeganga	62
5.3.2. Recuperación de imágenes utilizando CLIP con ajuste de <i>embeddings</i> visuales y <i>query</i> adaptativa	65
5.3.2.1. UNIQLO	66
5.3.2.2. Cartier	69
5.3.2.3. IKEA	73
5.3.2.4. World Market	76
5.3.2.5. Homy	80
5.3.2.6. Pepeganga	86
5.4. Sumario	93
6. Conclusiones	99

7. Trabajos futuros	101
Bibliografía	102
Anexo A. Anexo	104
A.1. Recuperación de imágenes utilizando UMAP y ajuste de <i>embeddings</i> visuales	104
A.1.1. <i>Embeddings</i> visuales sin ajustar	104
A.1.2. <i>Embeddings</i> visuales ajustados	105
A.2. Redes neuronales <i>feed-forward fully-connected</i> para transformación de <i>embeddings</i> visuales	106