

Contents

| | |
|--|------------|
| Resumen Ejecutivo | i |
| Abstract | iii |
| 1 Introduction | 1 |
| 1.1 Motivation | 1 |
| 1.2 Description and challenges | 2 |
| 1.3 Document outline | 3 |
| 2 Preliminaries | 5 |
| 2.1 Theoretical framework | 5 |
| 2.1.1 Convolutional neural networks (CNN) | 5 |
| 2.1.2 History of convolutional models in computer vision | 9 |
| 2.2 Literature review | 10 |
| 2.2.1 Deep learning for content-based image retrieval | 10 |
| 2.2.2 Deep learning for object detection | 11 |
| 2.2.3 RetinaNet | 12 |
| 2.2.4 Related work | 13 |
| 3 Methodology | 15 |
| 3.1 DocExplore dataset | 15 |

| | | |
|----------|--|-----------|
| 3.2 | Our deep proposal | 18 |
| 3.2.1 | Offline phase | 19 |
| 3.2.2 | Online phase | 25 |
| 4 | Results and Discussion | 30 |
| 4.1 | Evaluation | 30 |
| 4.2 | Text region filtering | 31 |
| 4.3 | Image retrieval and pattern spotting | 32 |
| 4.4 | Comparison with state-of-the-art | 42 |
| 5 | Conclusion | 46 |
| | Appendix | 48 |
| A | Cosine similarity | 48 |
| B | Coordinate translation | 48 |
| C | Query localization with multiple receptive field | 49 |
| D | mAP (mean Average Precision) | 50 |
| E | mAP relative improvement | 51 |
| F | 2×2 group assignation | 51 |
| G | ‘Square ornament’ and ‘Round ornament’ results | 51 |
| | Bibliography | 54 |