

Contents

	<i>page</i>
<i>Abstract</i>	<i>v</i>
<i>Acknowledgements</i>	<i>vii</i>
<i>Contents</i>	<i>ix</i>
<i>List of Figures</i>	<i>xi</i>
<i>List of Tables</i>	<i>xv</i>
<i>List of Symbols</i>	<i>xvii</i>
<i>List of Abbreviations</i>	<i>xix</i>
Chapter 1: Introduction	1
1.1 Motivation	1
1.2 Problem Statement	2
1.2.1 Overview of work done	3
1.3 Contribution	5
1.4 Publications	6
1.5 Organization	6
Chapter 2: Background and Related Work	9
2.1 Scene text in Classification	9
2.2 Scene text in Semantic Embedding	10
2.3 Scene text Datasets	12
2.4 Context and Attention	12
2.5 Scene text in Question Answering	15
2.6 Text Semantics	17
2.7 Conclusion	19
Chapter 3: Contextual Encoding	21
3.1 Overview	21
3.2 Method	22
3.2.1 Visual Encoder	22
3.2.2 Scene Text Encoder	23
3.2.3 Contextual Visual Text (VT) Encoder	23
3.3 Application of the model	26
3.3.1 Task 1: Image-Statement Relevance	26
3.3.2 Task 2: Classification	29
3.4 Experimental Results : Image-Statement Relevance	29
3.4.1 Metrics	30
3.4.2 Comparison with the state-of-the-art	30
3.4.3 Ablation Study	30
3.4.4 Qualitative Results	31
3.5 Experimental Results: Classification	33
3.5.1 Topic Classification	33
3.5.2 Tweet Classification	33
3.6 Computational Cost	35
3.7 Conclusion	35
Chapter 4: Knowledge enriched scene text	37
4.1 Overview	37
4.2 knowledge applied to Vision Language Understanding Tasks	38
4.2.1 Knowledge-enabled Datasets	38
4.2.2 External Knowledge applied to Vision Language	38
4.2.3 Using Knowledge to generate answer	40
4.3 EKTVQA: External knowledge-enabled Text-VQA	40
4.3.1 Knowledge Extraction And Validation	41
4.3.2 Multimodal Features	43
4.3.3 Reasoning with External Knowledge	44
4.4 Experiments and Results	47
4.4.1 Implementation and Dataset details	47
4.4.2 Comparison with state-of-the-art	49

4.4.3	How knowledge <i>helps</i>	51
4.4.4	Ablation Studies	54
4.4.5	<i>Bottleneck and Limitations</i>	56
4.4.6	Results Summary	57
4.5	Conclusion	57
Chapter 5: Bias due to scene text		59
5.1	Introduction	59
5.2	Scene text Reference	60
5.3	Scene text Dependence	64
5.4	Conclusion	66
Chapter 6: Conclusions		67
6.1	Summary of proposed methods and Contributions	67
6.2	Summary of Experimental Results and Insights	68
6.3	Future Work	69
Bibliography		73