CROSS – MODAL BIDIRECTIONAL TRANSLATION VIA REINFORCEMENT LEARNING

2020. 06. 10.
Moon Sungwon

INTRODUCTION

TEXT

이 영화 진짜 재미 없다

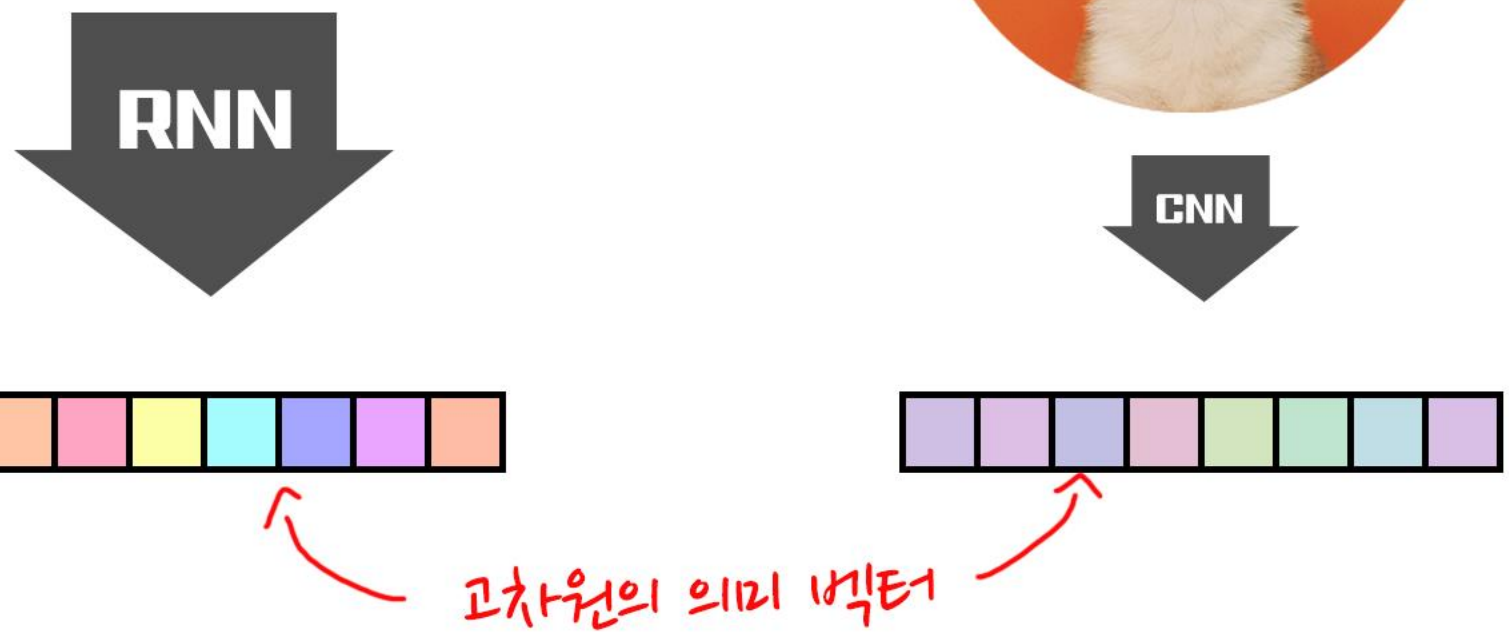
Class: 부정 (99%)

Image

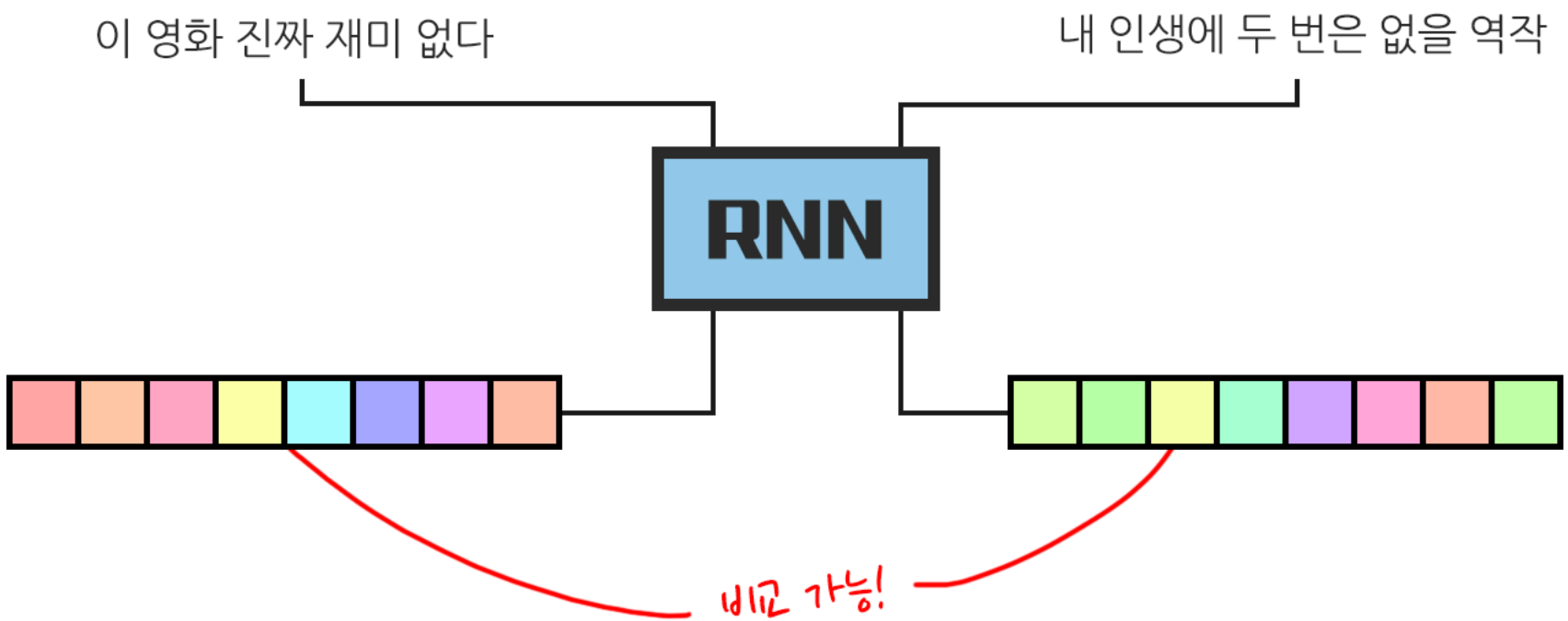
Class: 강아지 (87%)

INTRODUCTION

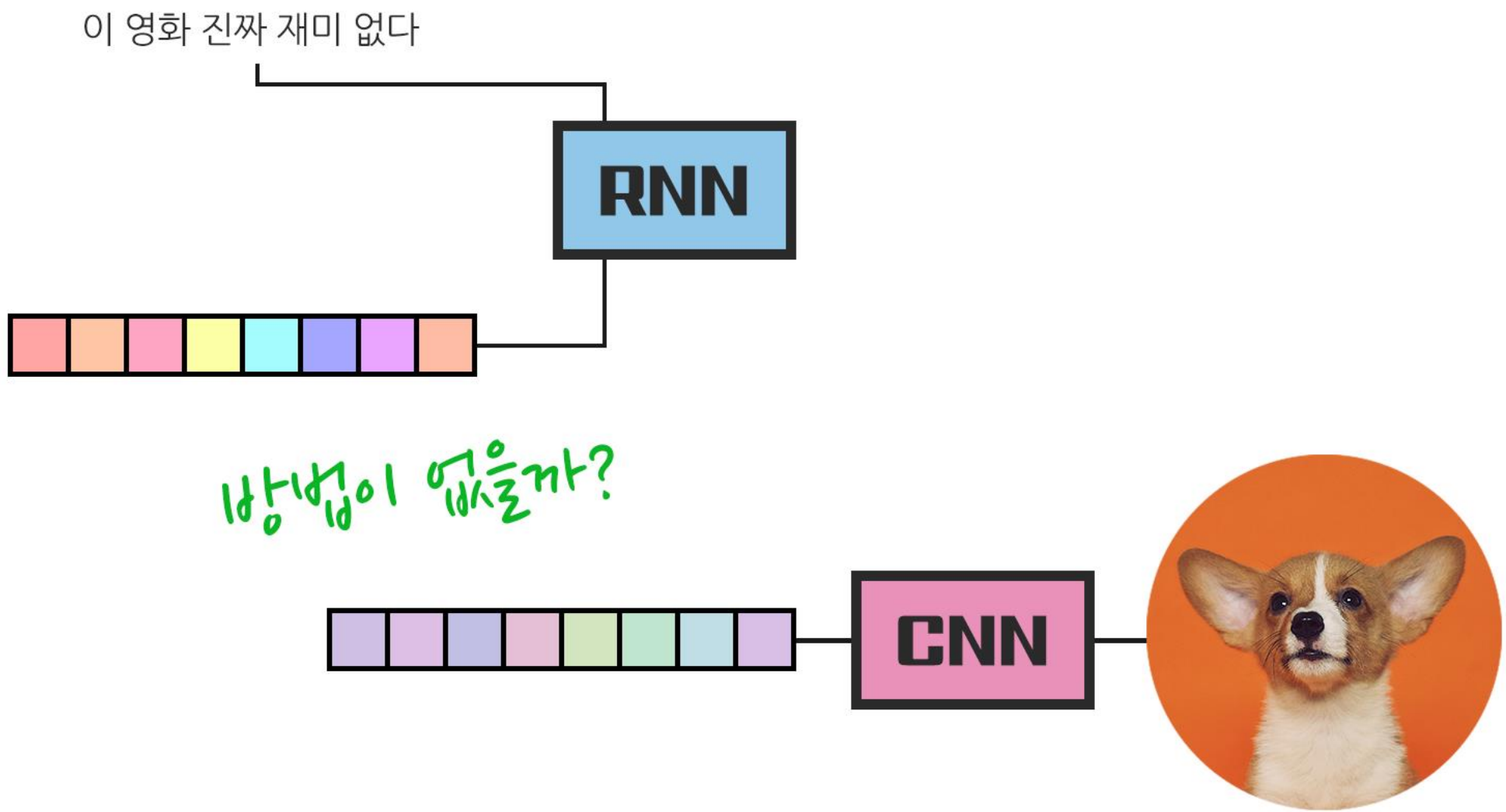
이 영화 진짜 재미 없다



INTRODUCTION



INTRODUCTION



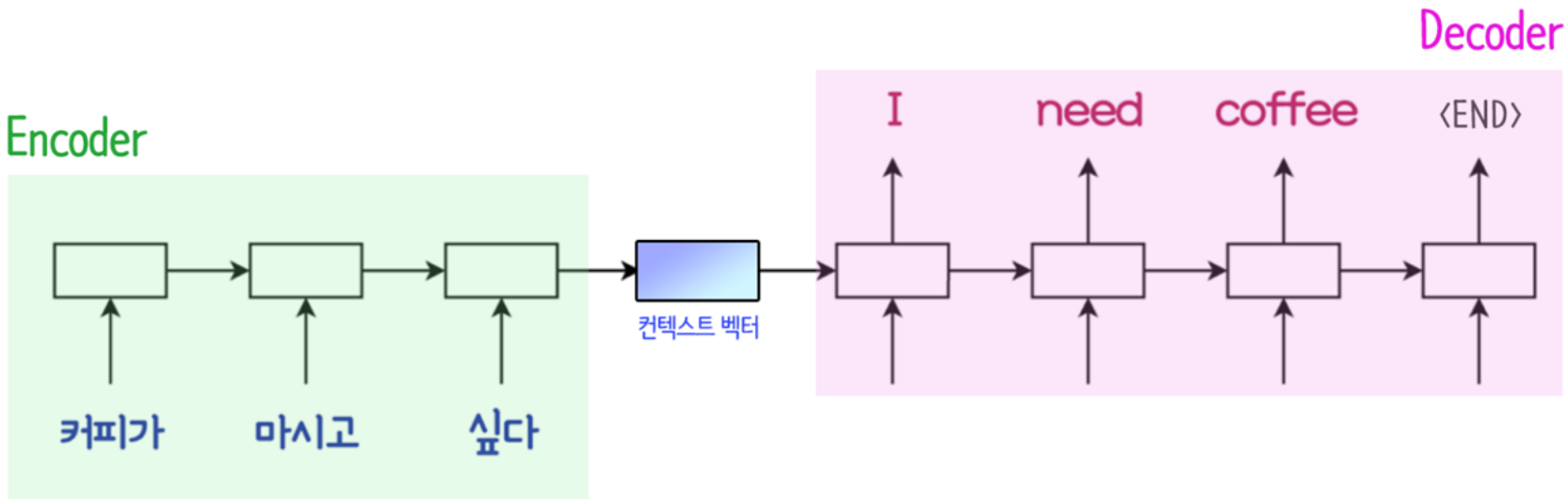
INTRODUCTION

“ **Cross-modal Translation** ”

= Seq2Seq

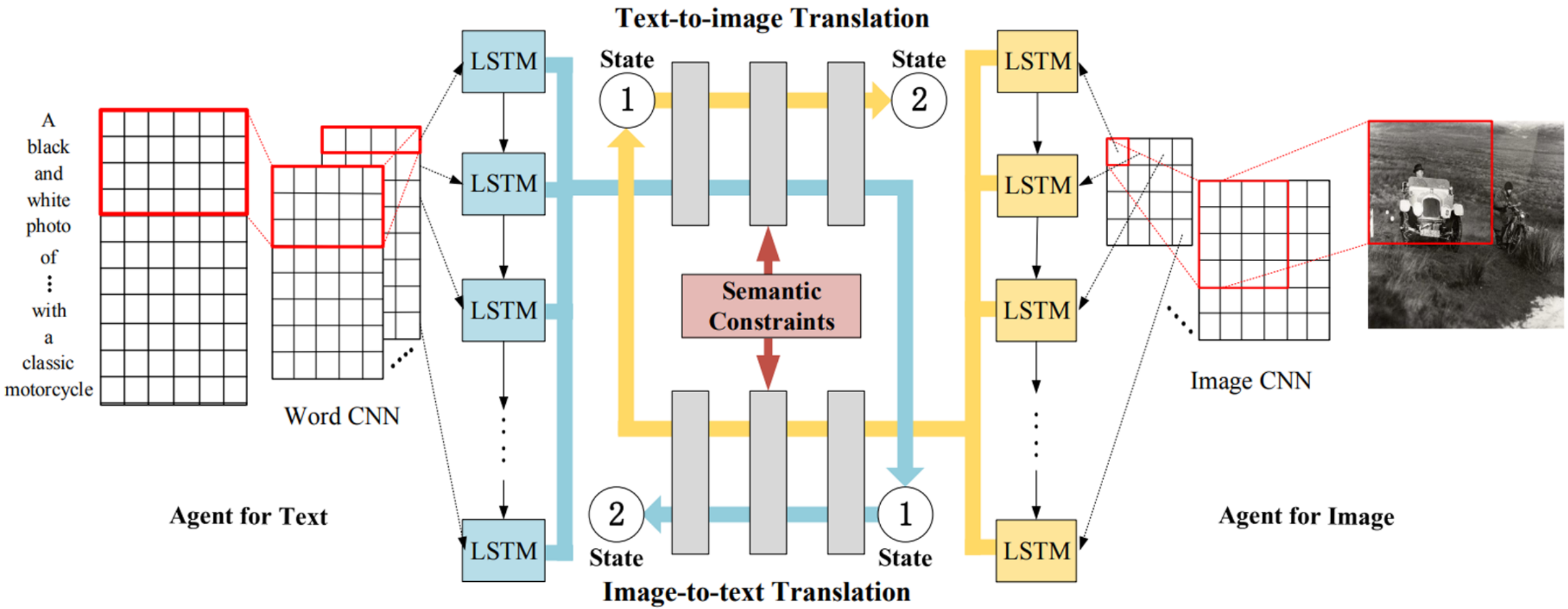
INTRODUCTION

기존의 방법이 대체 어땠길래?

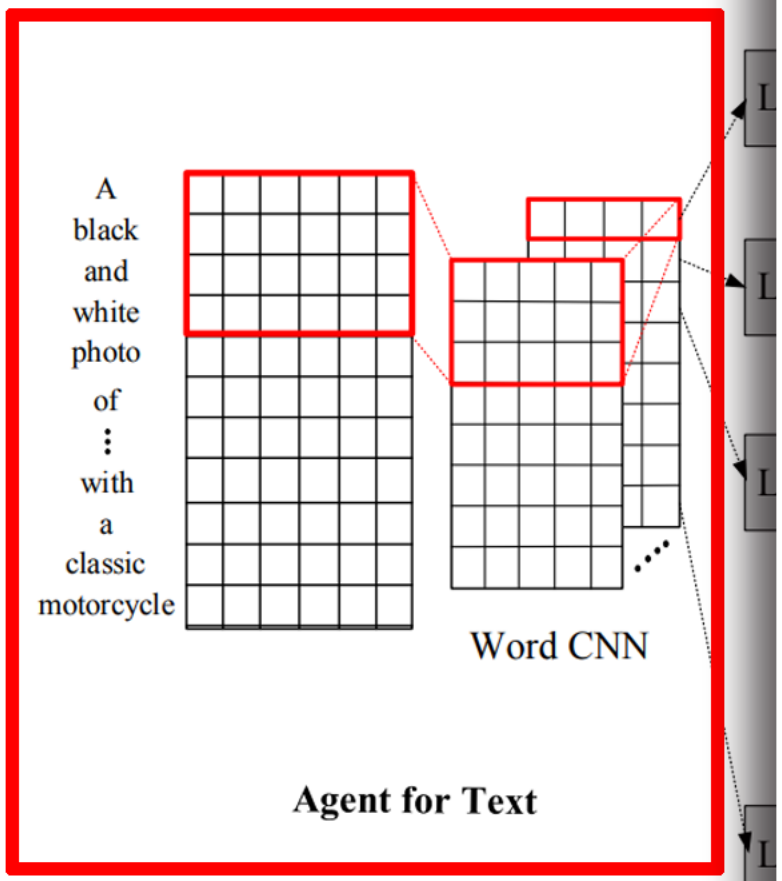


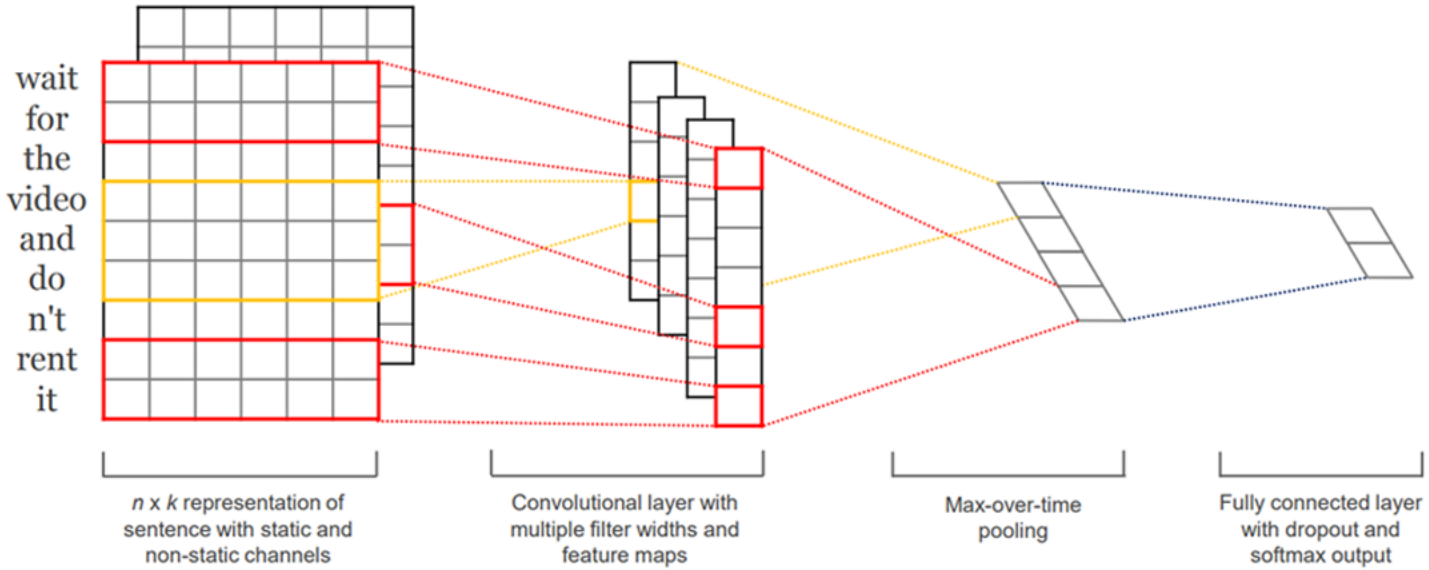
▲ Sequence-to-Sequence (Seq2seq)

MODEL


Text-to-image Translation
A black and white photo of ⋮ with a classic motorcycle
Word CNN
LSTM
LSTM
LSTM
LSTM
State 1
State 2
Semantic Constraints
2 State
1 State
Image-to-text Translation
LSTM
LSTM
LSTM
LSTM
Image CNN
Agent for Text
Agent for Image

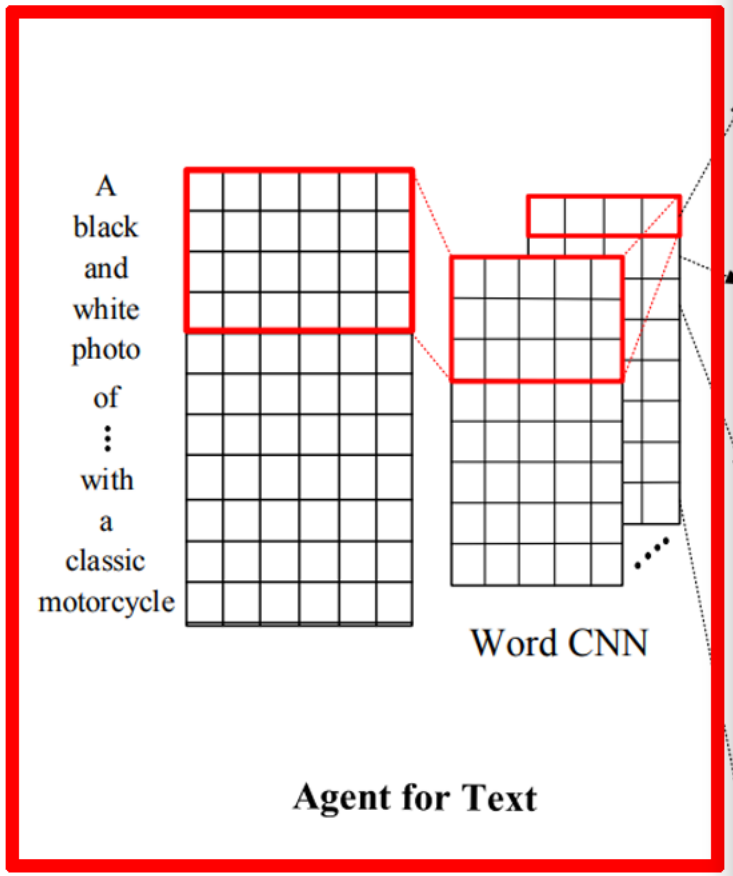
MODEL

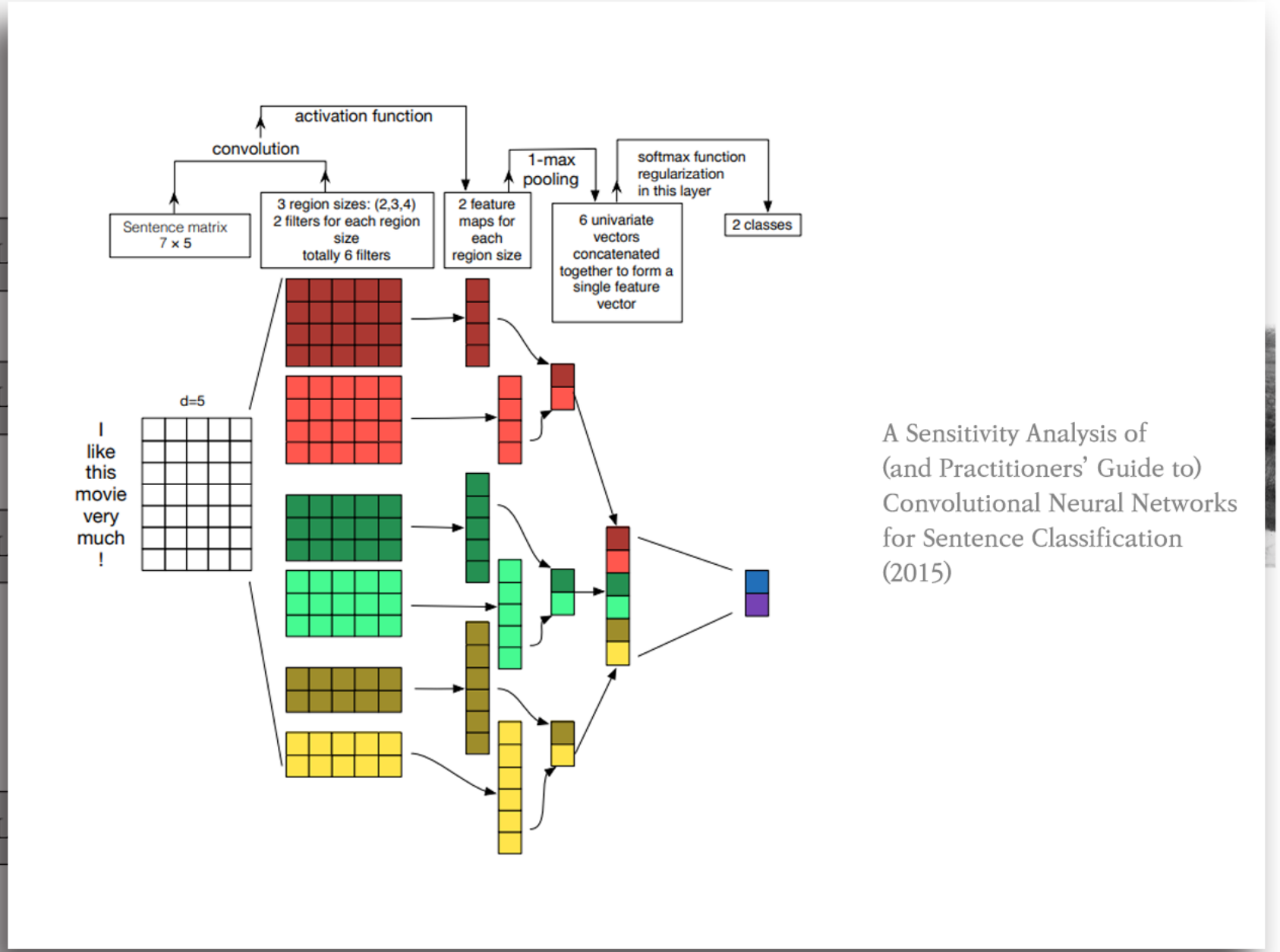




Convolutional Neural Networks for Sentence Classification (2014)

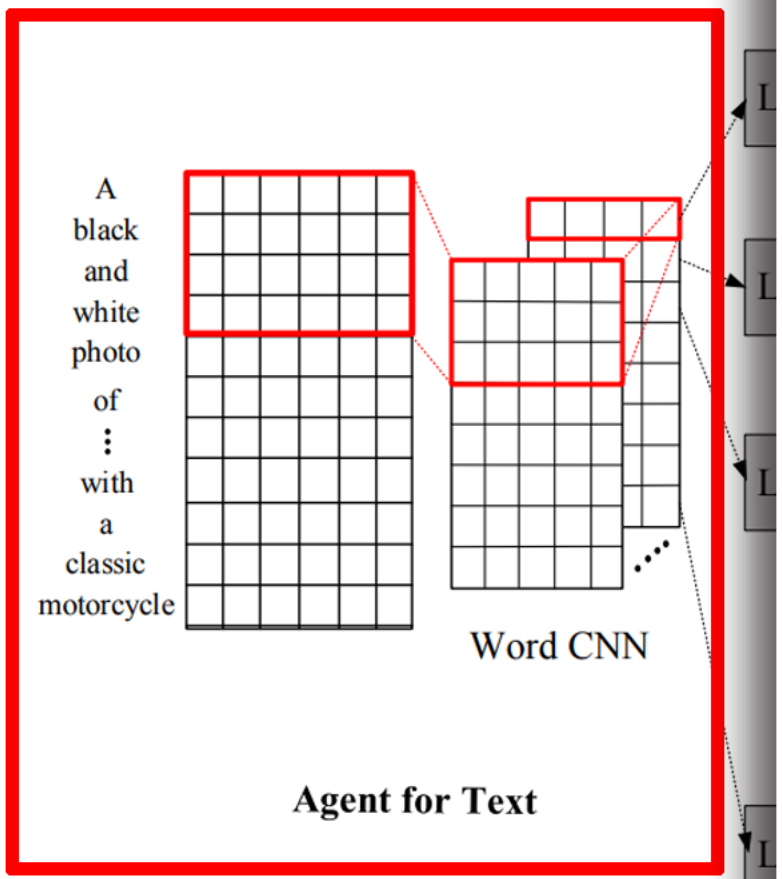
MODEL

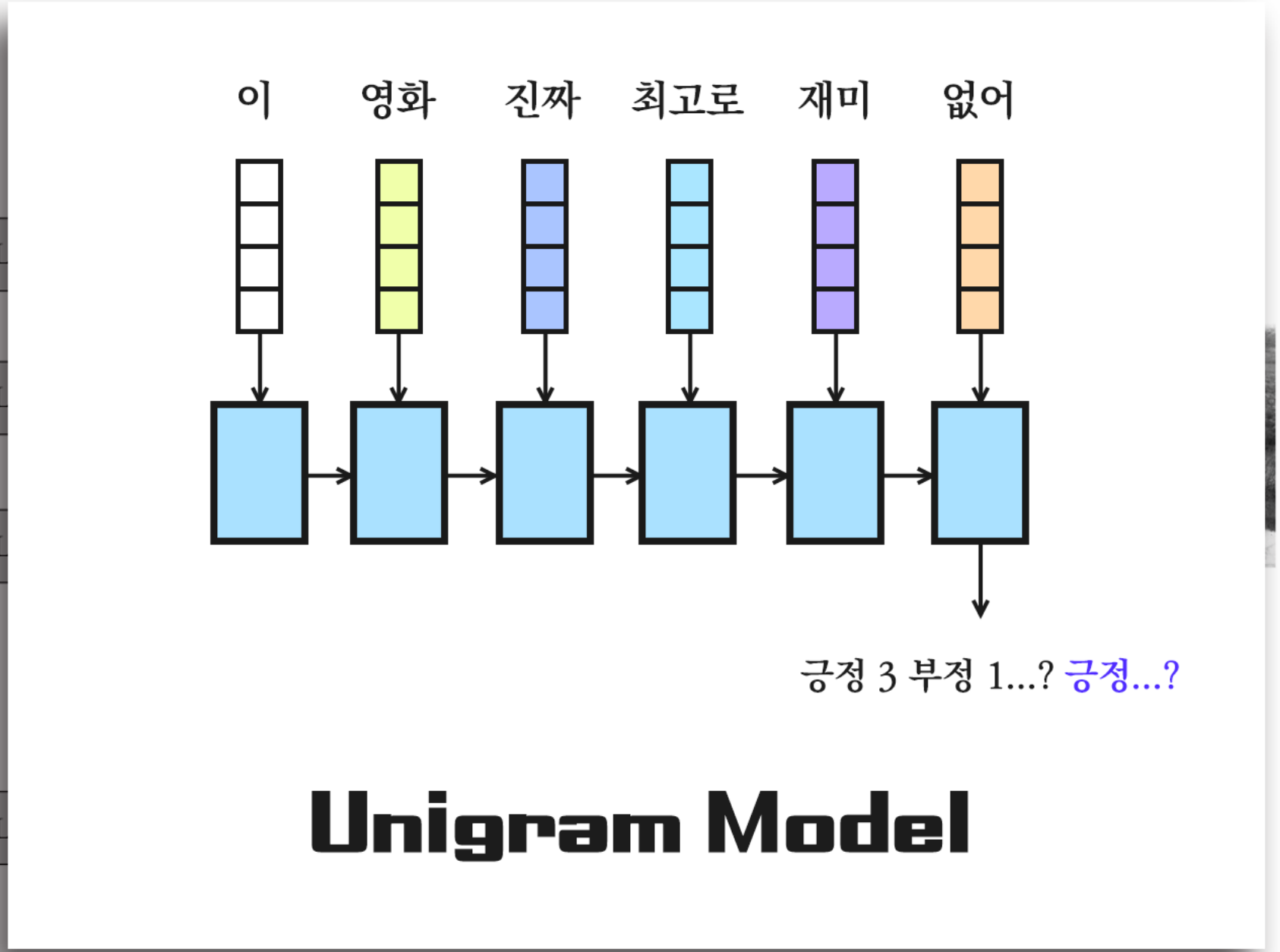




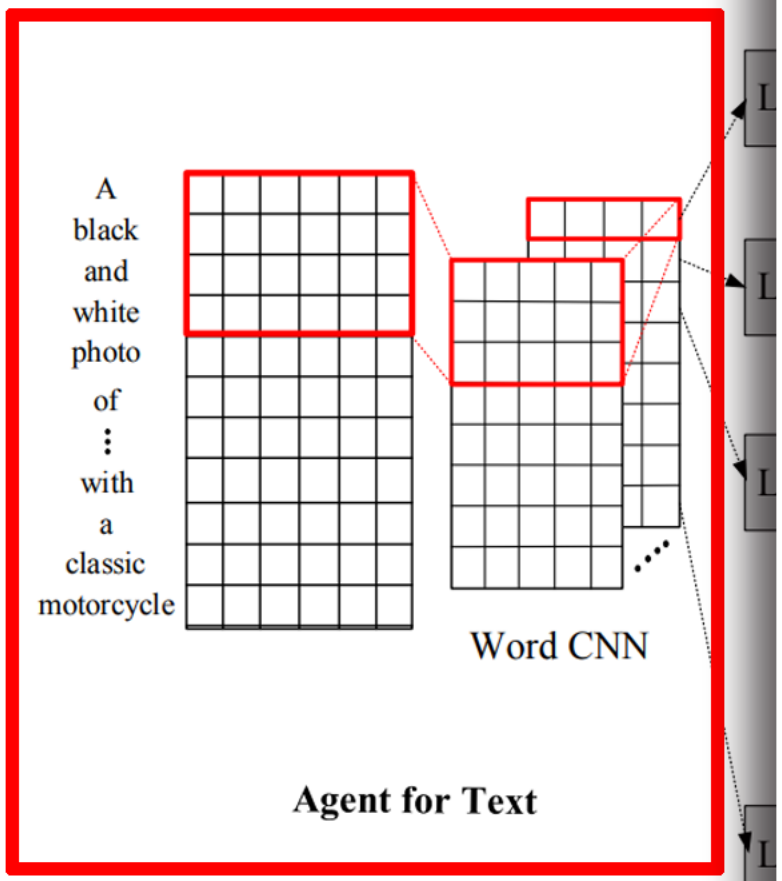
A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification (2015)

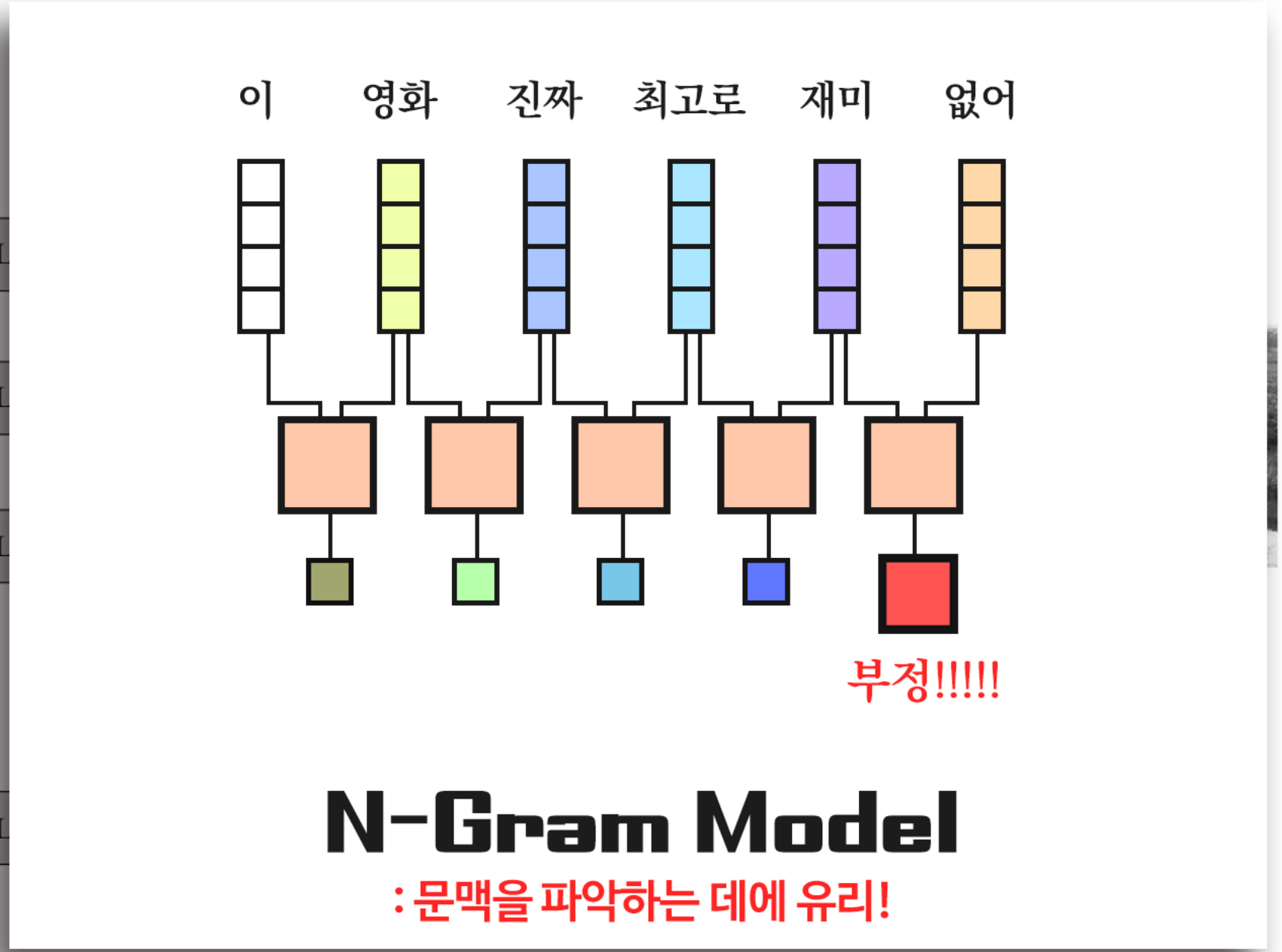
MODEL


A
black
and
white
photo
of
with
a
classic
motorcycle
Word CNN
Agent for Text

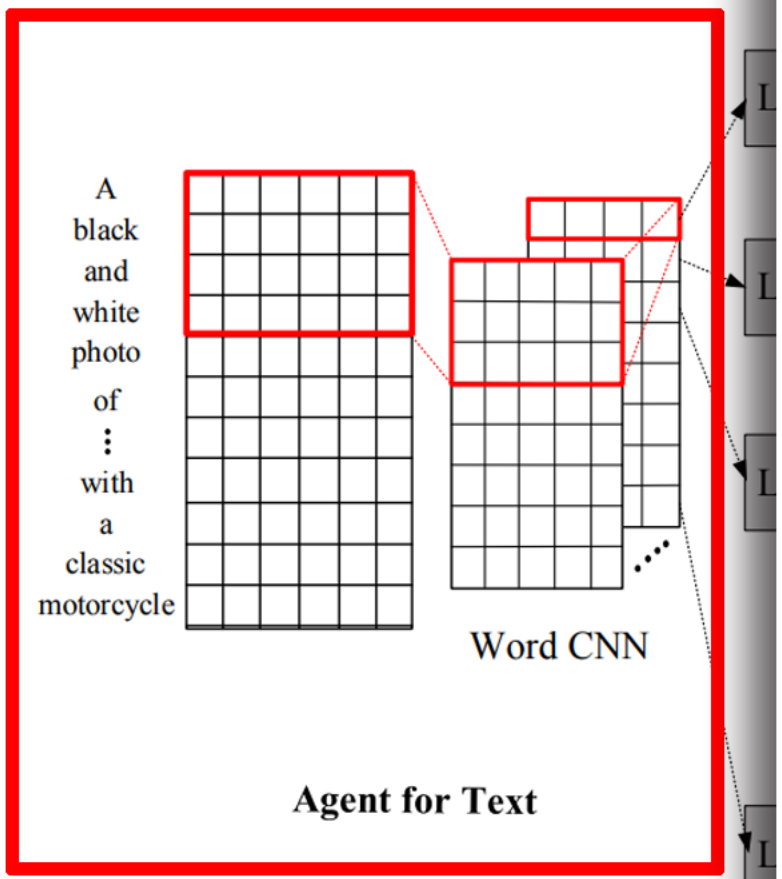

이
영화
진짜
최고로
재미
없어
긍정 3 부정 1...? 긍정...?
Unigram Model

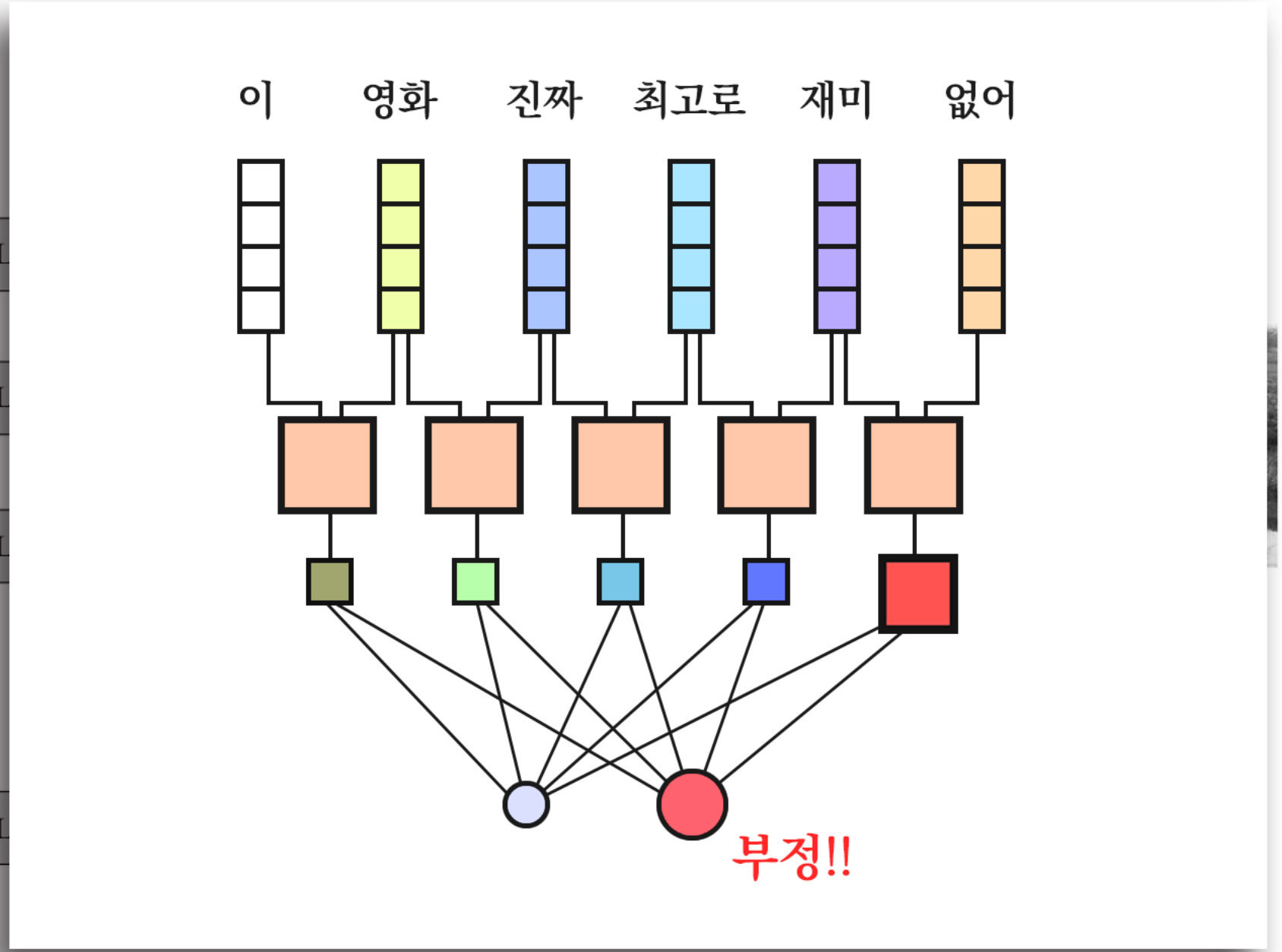
MODEL


A
black
and
white
photo
of
with
a
classic
motorcycle
Word CNN
Agent for Text

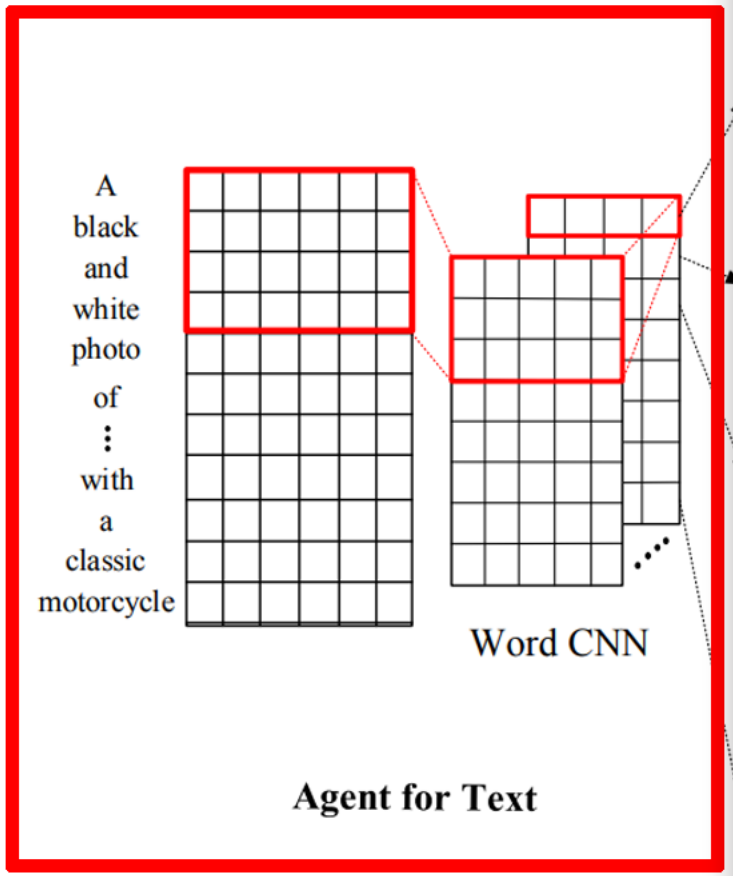

이
영화
진짜
최고로
재미
없어
부정!!!!!
N-Gram Model
: 문맥을 파악하는 데에 유리!

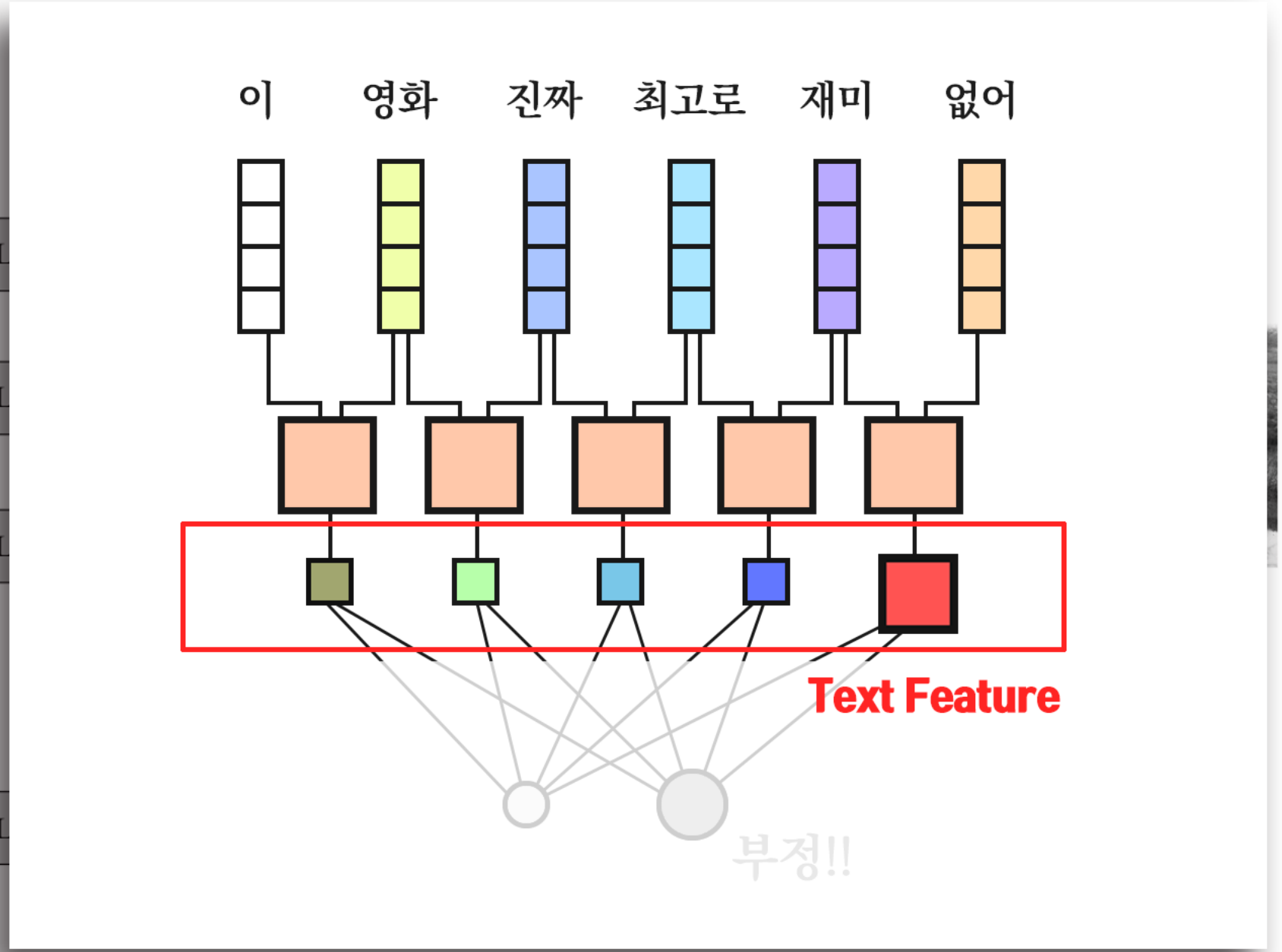
MODEL


A
black
and
white
photo
of
with
a
classic
motorcycle
Word CNN
Agent for Text

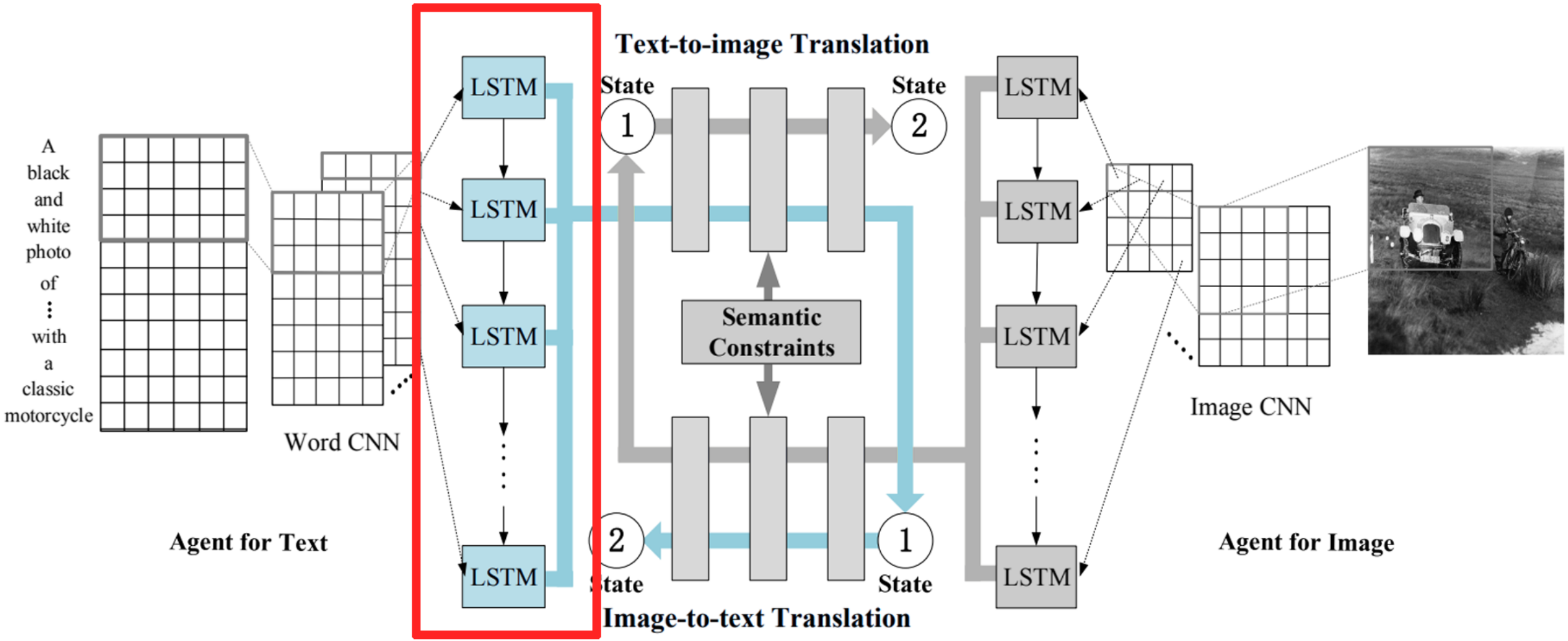

이
영화
진짜
최고로
재미
없어
부정!!

MODEL

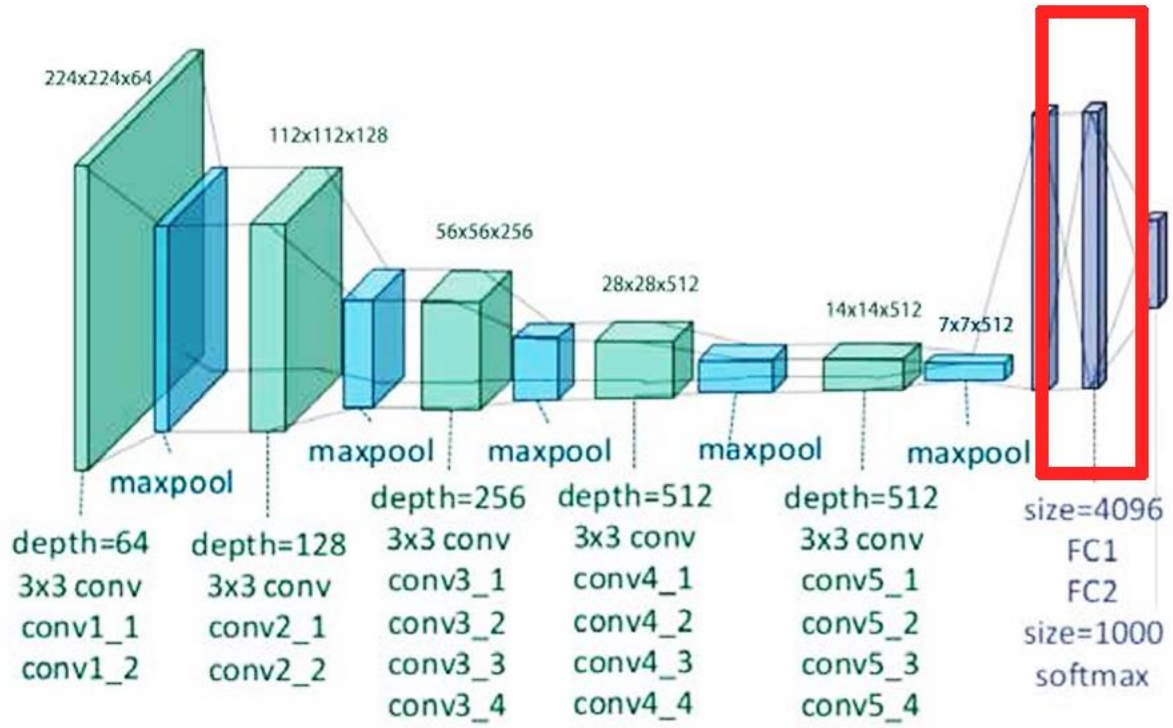

A
black
and
white
photo
of
with
a
classic
motorcycle
Word CNN
Agent for Text


이
영화
진짜
최고로
재미
없어
Text Feature
부정!!

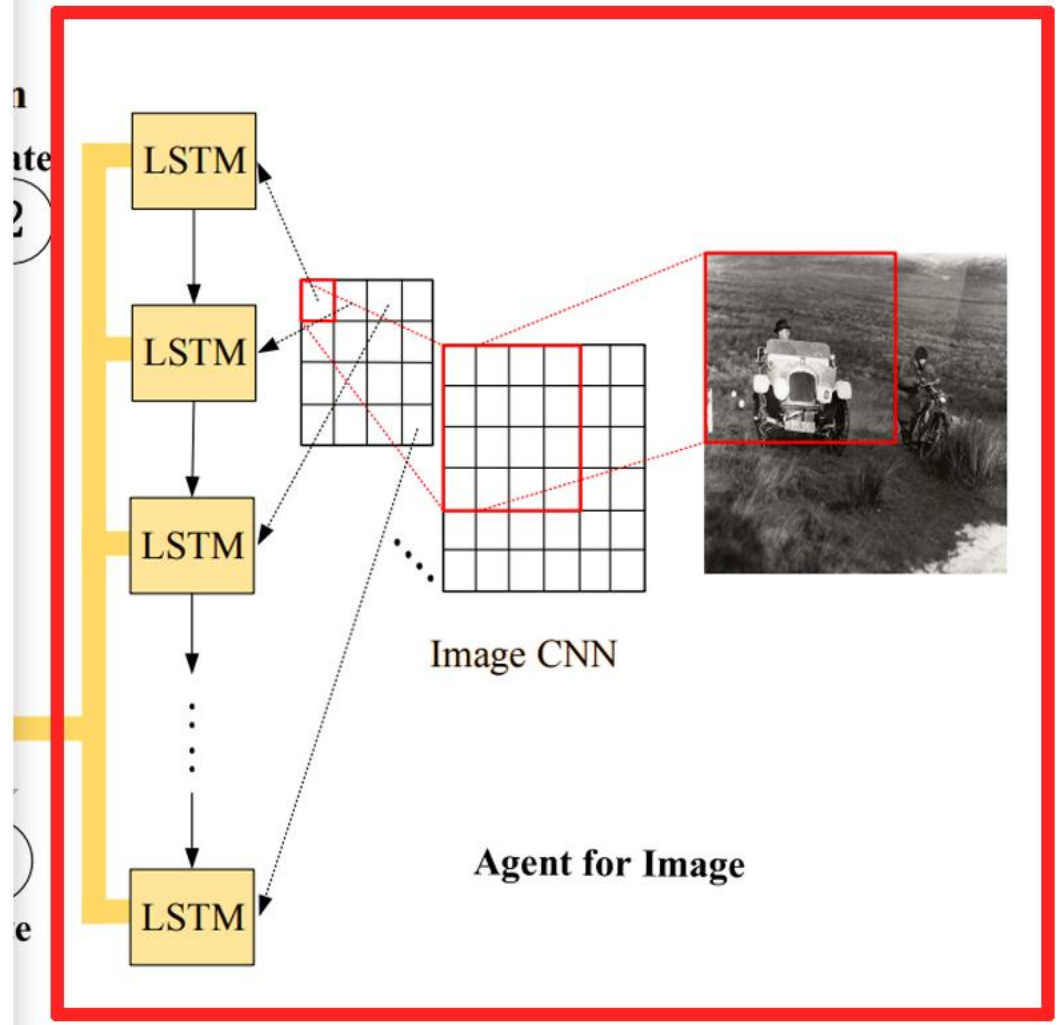
MODEL


A black and white photo of ⋮ with a classic motorcycle
Word CNN
Agent for Text
LSTM
LSTM
LSTM
LSTM
Text-to-image Translation
State 1
State 2
Semantic Constraints
State 2
State 1
Image-to-text Translation
LSTM
LSTM
LSTM
LSTM
Image CNN
Agent for Image

MODEL

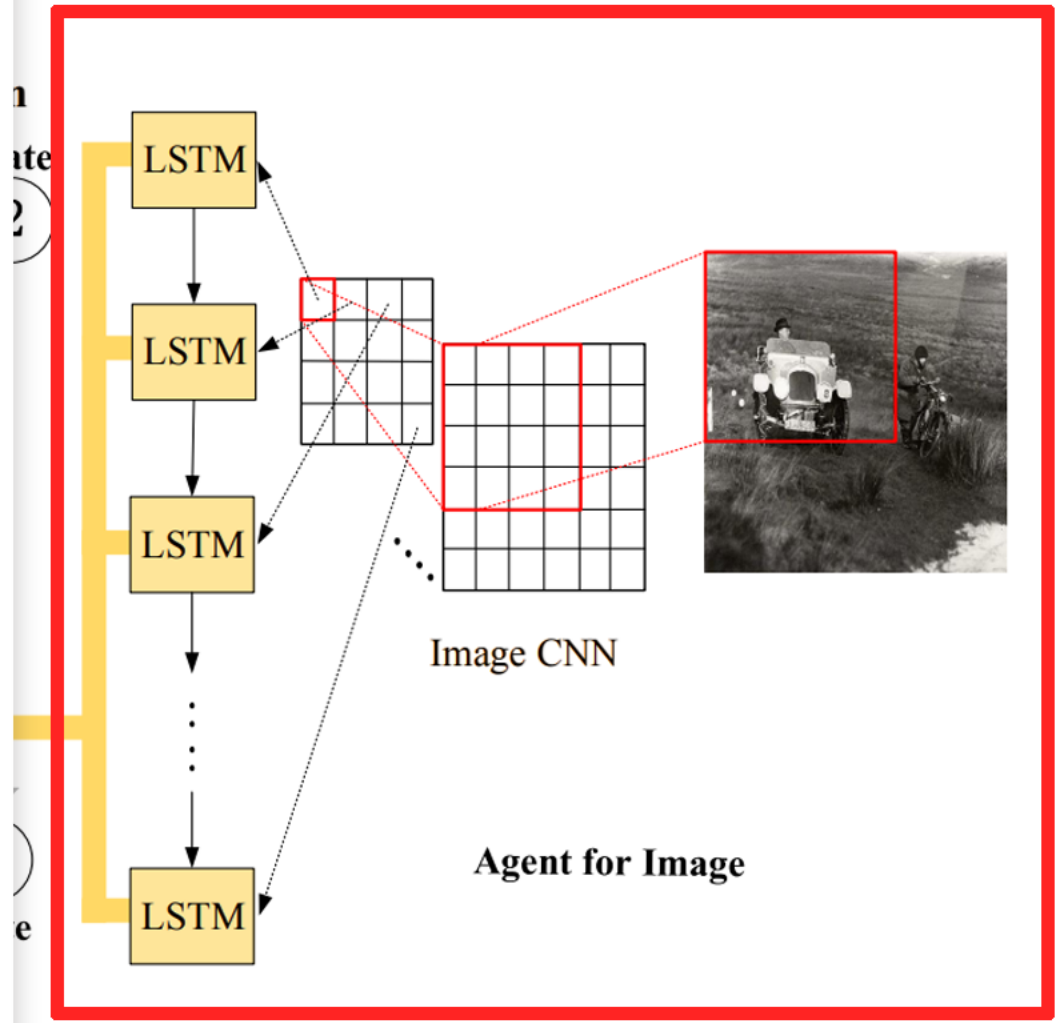


VERY DEEP CONVOLUTIONAL NETWORKS
FOR LARGE-SCALE IMAGE RECOGNITION
(2014)



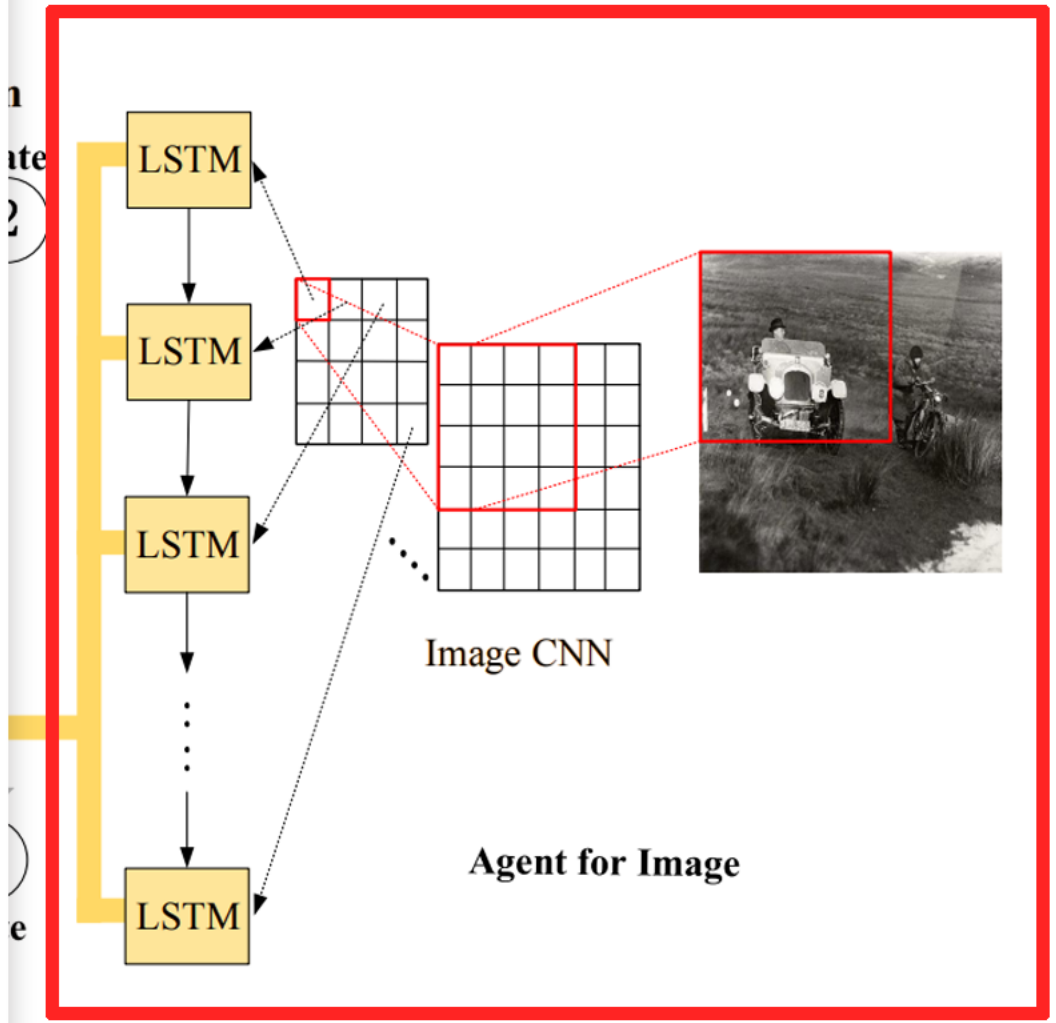
MODEL

VGG-19 Input size



MODEL


Feature 1
Feature 3
Feature 2
Feature 4


LSTM
LSTM
LSTM
LSTM
Image CNN
Agent for Image

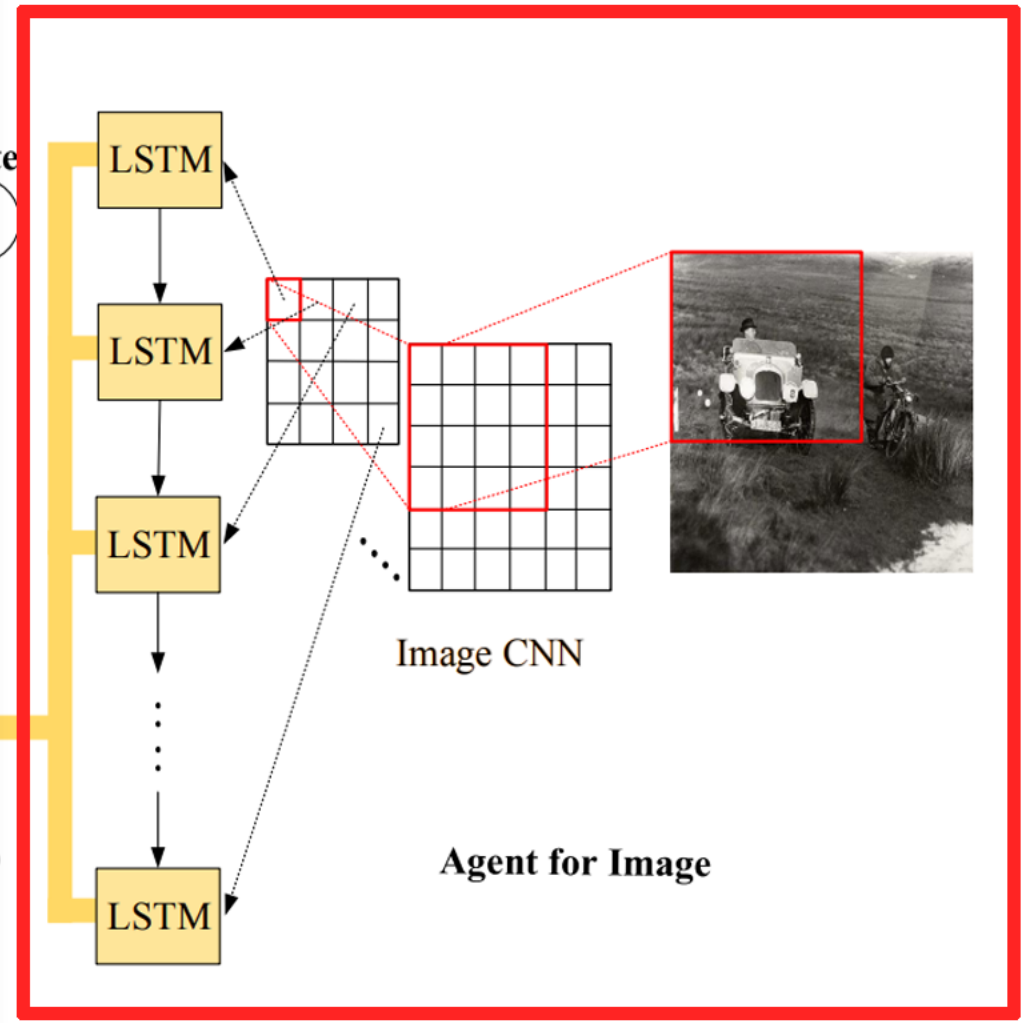
MODEL

Feature Map

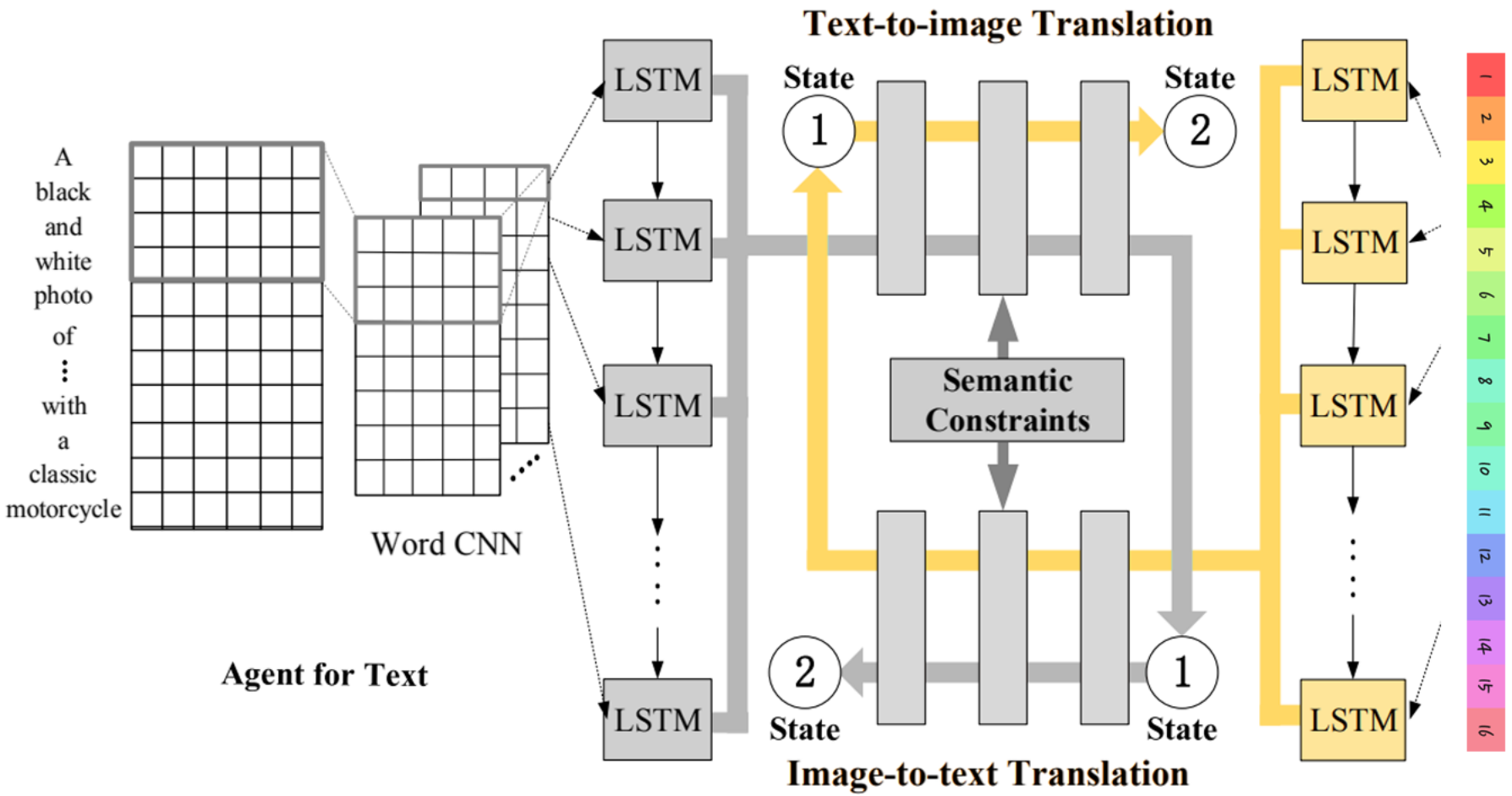
1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	16

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

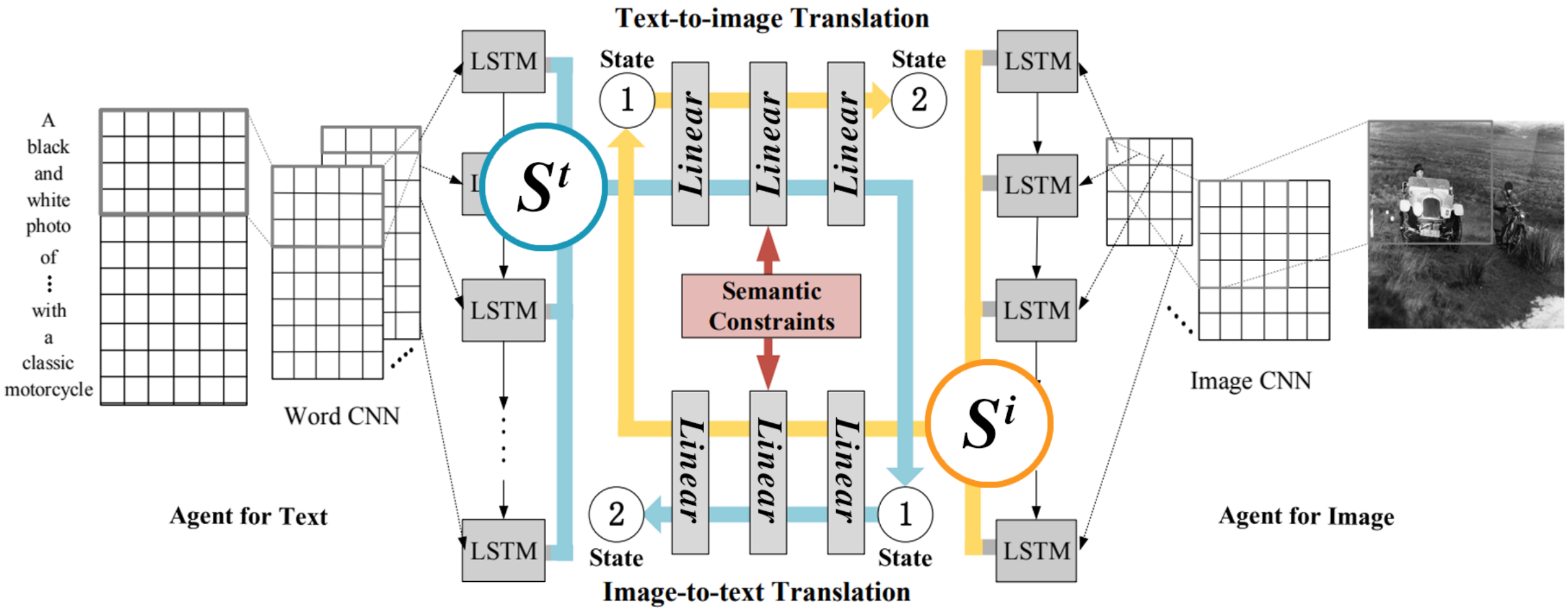
시선의 이동 →



MODEL


Text-to-image Translation
A black and white photo of ⋮ with a classic motorcycle
Word CNN
Agent for Text
LSTM
LSTM
LSTM
LSTM
State 1
State 2
Semantic Constraints
State 2
State 1
LSTM
LSTM
LSTM
LSTM
Image-to-text Translation

MODEL

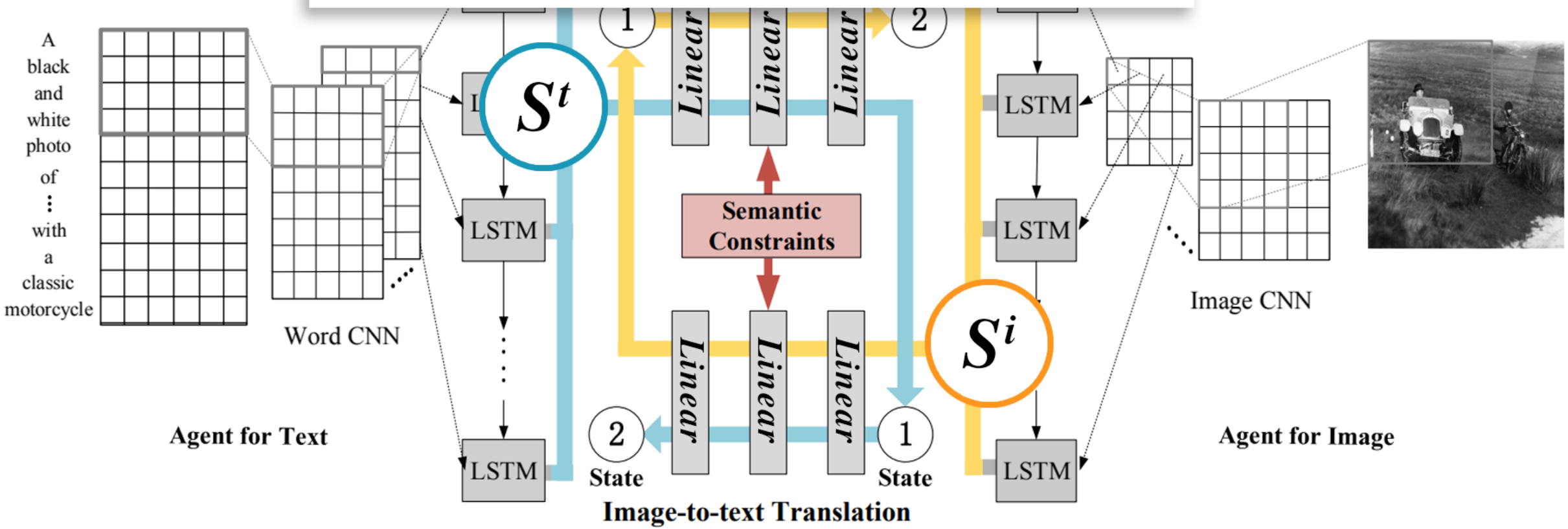

Text-to-image Translation
A black and white photo of ⋮ with a classic motorcycle
Word CNN
LSTM
LSTM
LSTM
LSTM
S^t
State 1
Linear
Linear
Linear
State 2
Semantic Constraints
Linear
Linear
Linear
State 2
State 1
S^i
LSTM
LSTM
LSTM
LSTM
Image CNN
Agent for Text
Agent for Image
Image-to-text Translation

MODEL

State: S^t

Action: *Linear*

$\rightarrow S^i_{mid} \leftrightarrow S^i \rightarrow Reward^{inter}$



MODEL

State: S^i_{mid}

Action: *Linear*

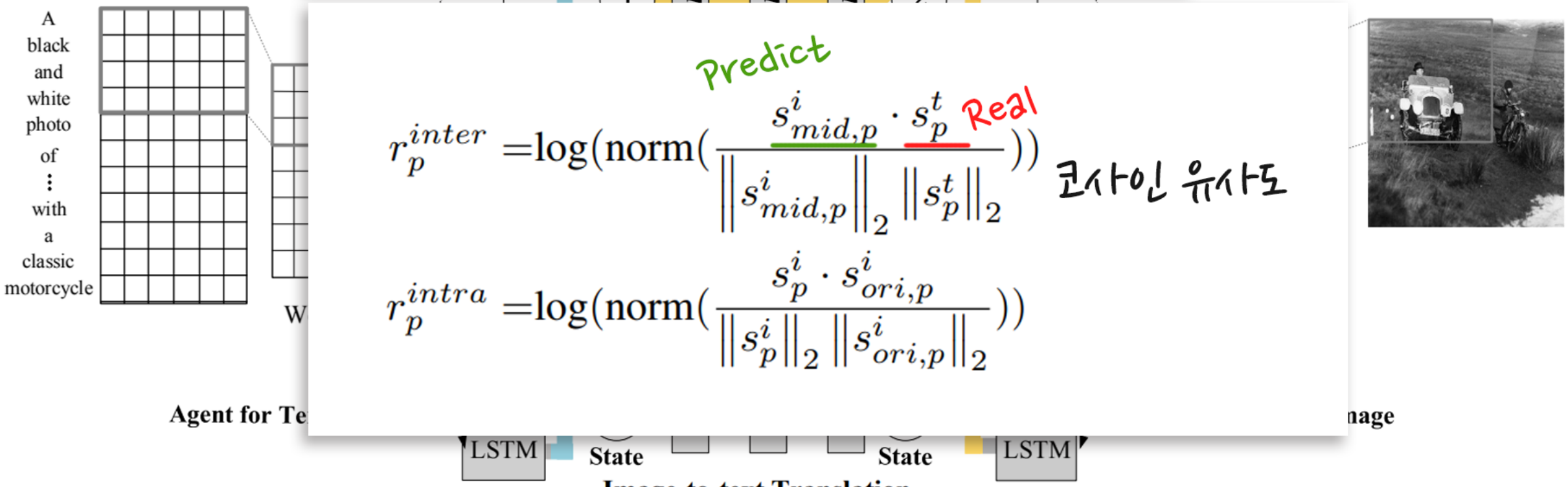
$\rightarrow S^t_{ori} \leftrightarrow S^t \rightarrow Reward^{intra}$

Predict

$$r_p^{inter} = \log(\text{norm}(\frac{s^i_{mid,p} \cdot s^t_p}{\left\|s^i_{mid,p}\right\|_2 \left\|s^t_p\right\|_2})) \quad \text{Real}$$

코사인 유사도

$$r_p^{intra} = \log(\text{norm}(\frac{s^i_p \cdot s^i_{ori,p}}{\left\|s^i_p\right\|_2 \left\|s^i_{ori,p}\right\|_2}))$$

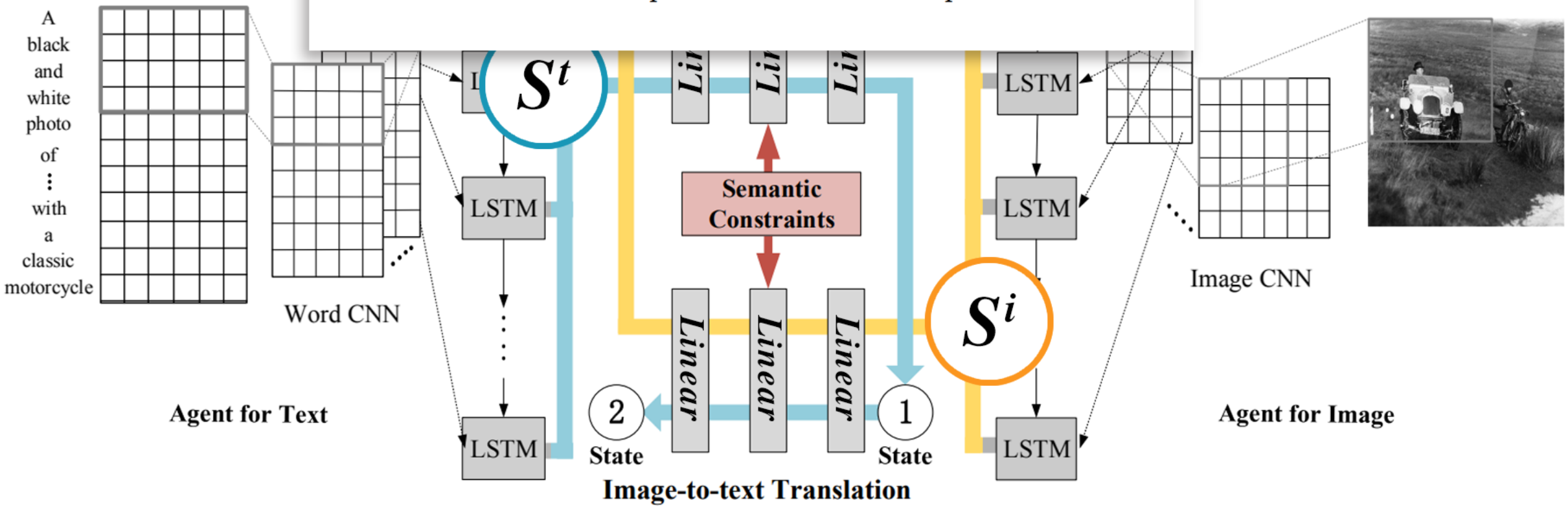


MODEL

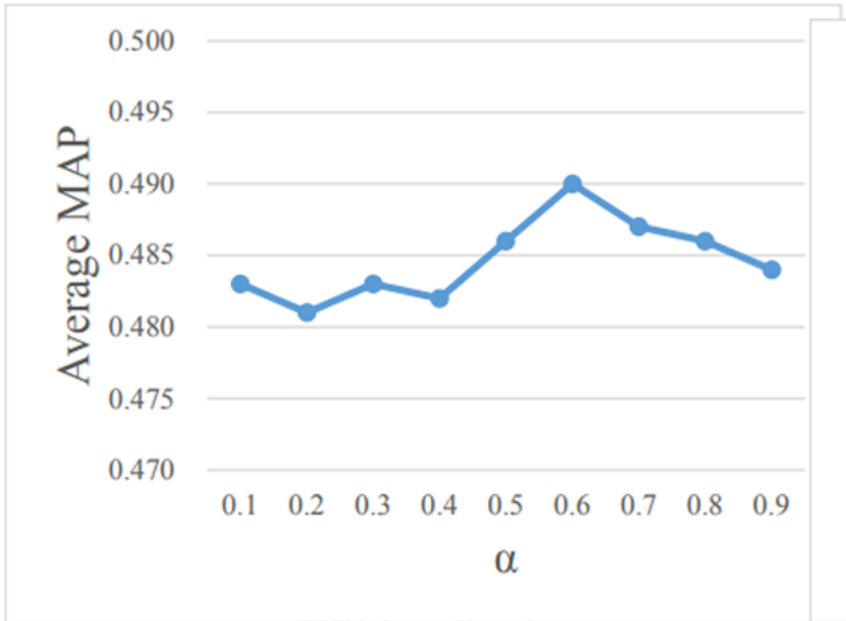
Text → Image → Text

R^{inter} R^{intra}

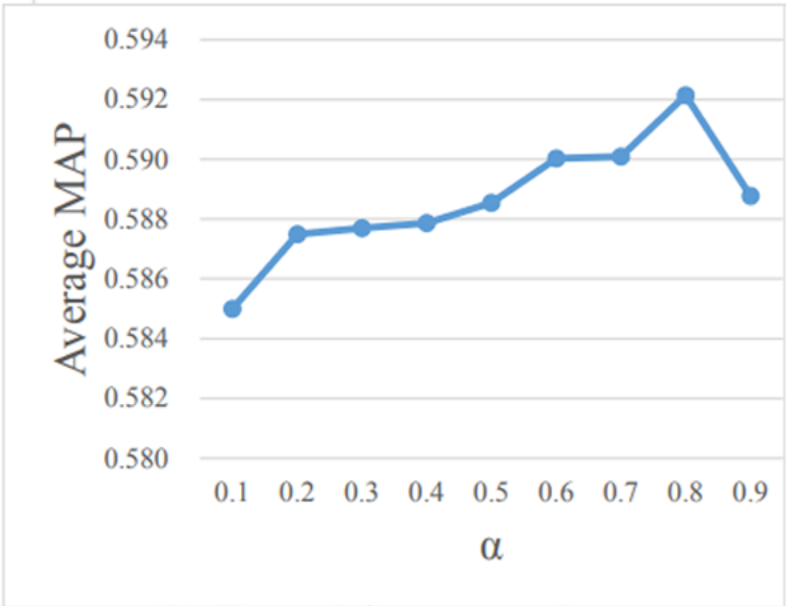
$$r_p = \alpha r_p^{inter} + (1-\alpha) r_p^{intra}$$



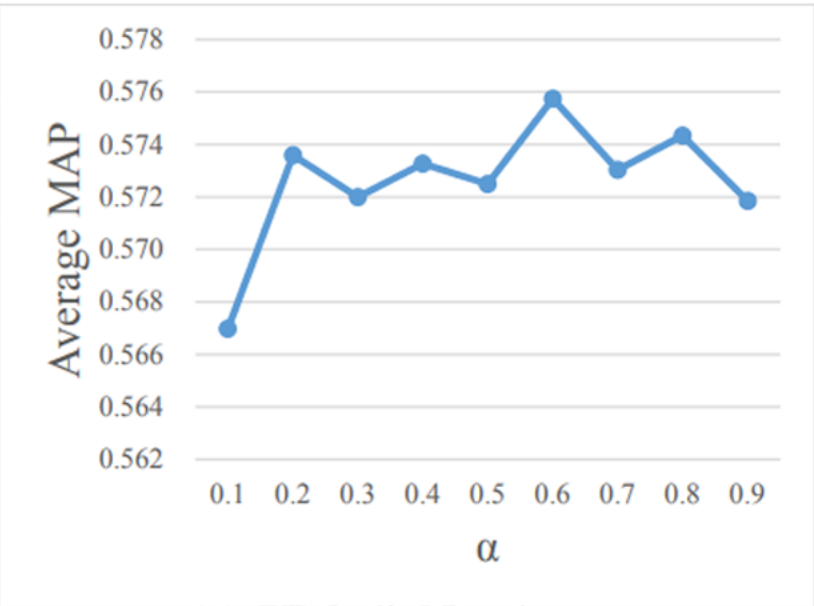
MODEL



(a) Wikipedia dataset



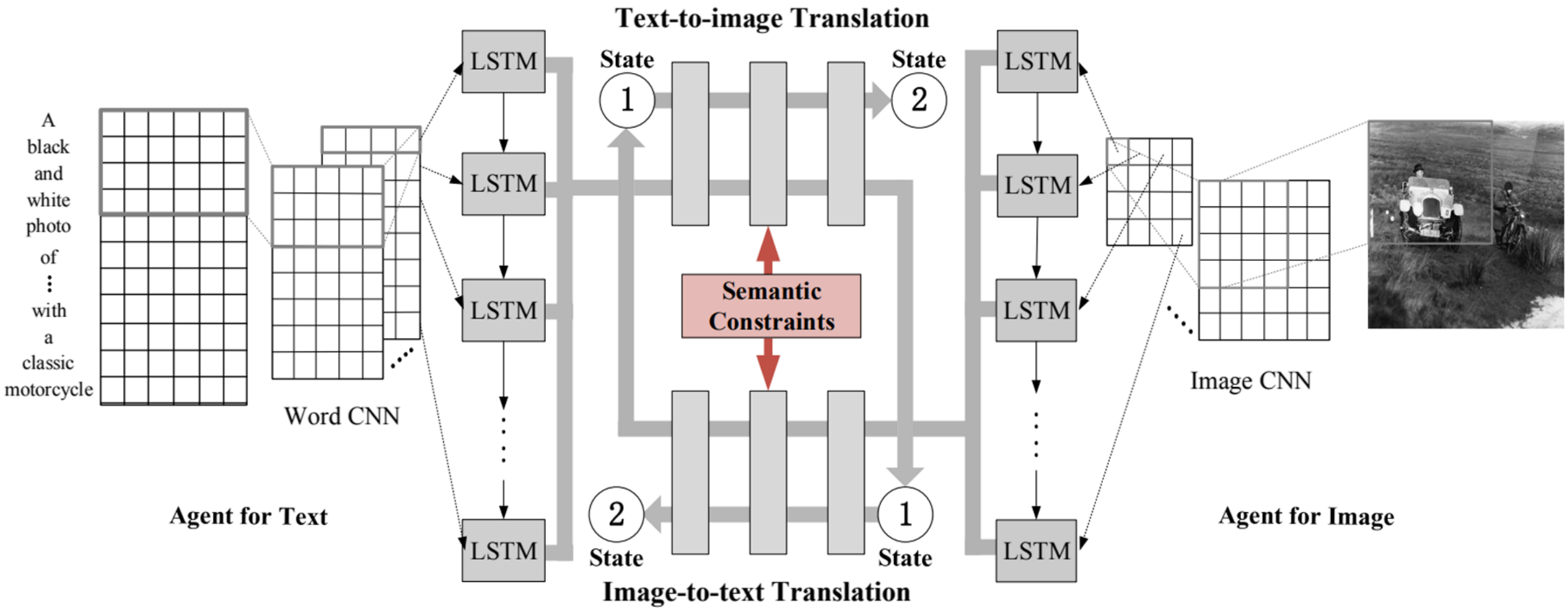
(b) Pascal Sentence dataset



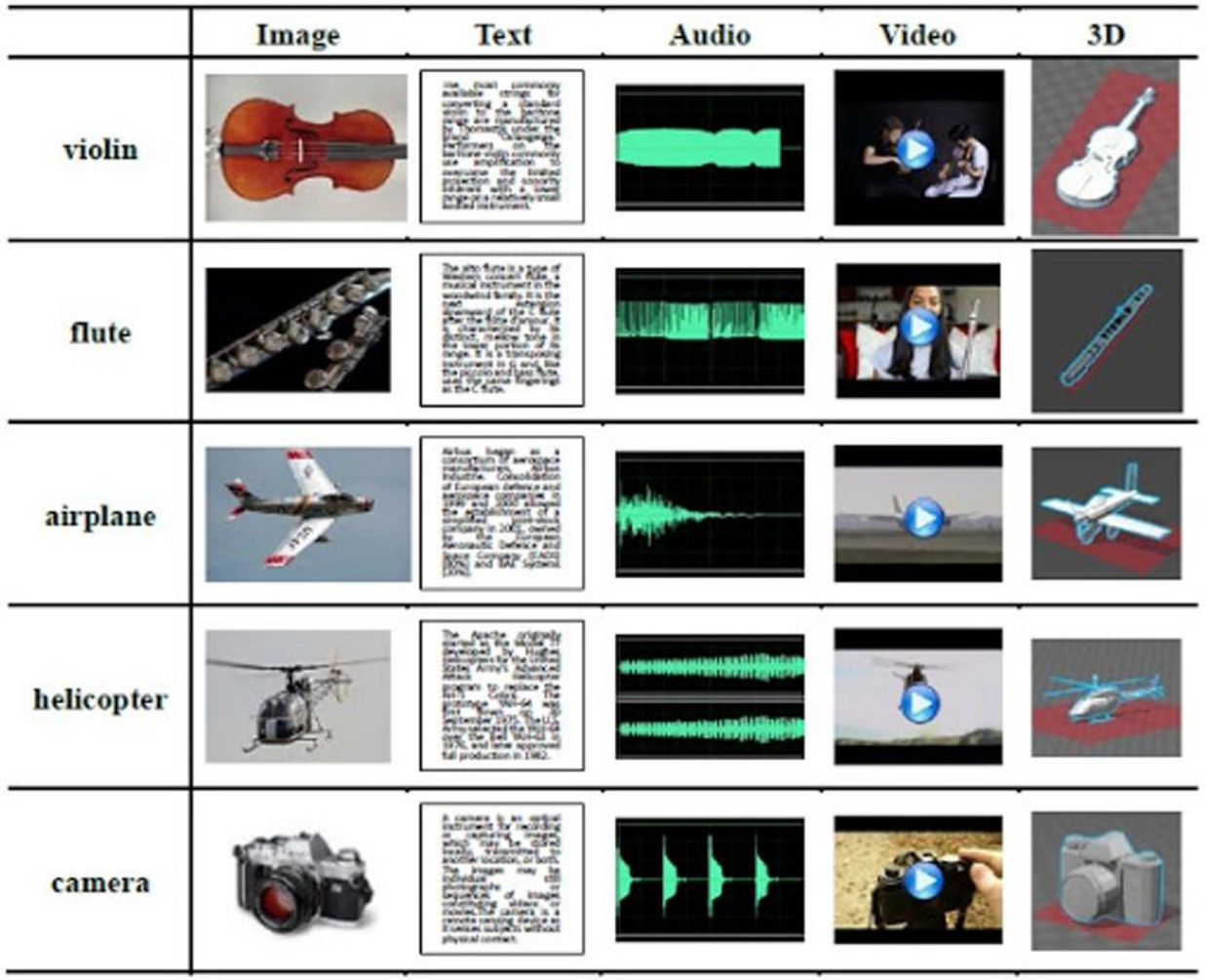
(c) XMediaNet dataset

$$r_p = \alpha r_p^{inter} + (1 - \alpha) r_p^{intra}$$

MODEL


Text-to-image Translation
A black and white photo of ⋮ with a classic motorcycle
Word CNN
LSTM
LSTM
LSTM
LSTM
State 1
State 2
Semantic Constraints
State 2
State 1
LSTM
LSTM
LSTM
LSTM
Image CNN
Agent for Text
Agent for Image
Image-to-text Translation

RESULTS



▲ PKU XMediaNet Dataset

Method	MAP scores		
	Image→Text	Text→Image	Average
Our CBT Approach	**0.577**	**0.575**	**0.576**
CCL	0.537	0.528	0.533
ACMR	0.536	0.519	0.528
CMDN	0.485	0.516	0.501
Deep-SM	0.399	0.342	0.371
LGCFL	0.441	0.509	0.475
JRL	0.488	0.405	0.447
DCCA	0.425	0.433	0.429
Corr-AE	0.469	0.507	0.488
KCCA	0.252	0.270	0.261
CFA	0.252	0.400	0.326
CCA	0.212	0.217	0.215

- 끝!