

모델의 탄생과 죽음

[OCR 5조] ConnectNet

T1097 서준배

T1138 윤준호

T1201 조호성

T1086 배철환

T1239 임기홍

T1046 김현우



서준배 (발표자)
#SATRN 분석
#하이퍼파라미터
실험



윤준호
#논문섭렵
#모델,
Augmentation



배철환
#구현담당
#CSTR, SRN 실험



조호성
#논문섭렵
#모델수정



김현우
#모델 서빙
#EDA 분석



임기홍
#데이터셋 추가
#데이터셋 분석

1. 문제 정의와 계획

2. 실험 성과

Dataset

Model Architecture

Hyperparameter Tuning & Ensemble

3. 모델의 탄생과 죽음

SRN

CSTR

실패로부터 배운 것들

모델 시연 

1. 문제 정의와 계획

주어진 Task

Math expression recognition

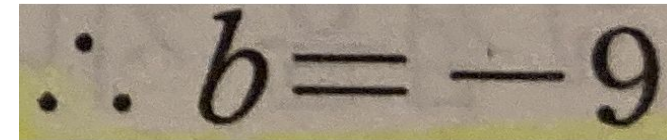
유사한 Task

- Image to text translation
- Multi-line OCR (Optical Character Recognition)

레퍼런스로 선택한 논문

- SATRN (Baseline)
- Aster
- SRN
- CSTR

Image to LaTeX



`\therefore b = - 9`

1. 문제 정의와 계획

회의록 (HackMD)

HackMD

Choose your favorite editor theme!

6 ONLINE

B I S H

19 - beam search를 적용할 수 있을까?

20 - 검색해보니 RNN 계열만 beam search를 쓰던데, CSTR은 CNN 계열이라 못쓸까??

21 - 다른 걸 적용할 수 있을까?

22 - 그나마 hyperparameter 조정??

23 - 아니면 단계별 학습??

24 - 윤준호

25 - CNN depth 32, cosine annealing, augmentations 모델 조금씩 튜닝하면서 계속 학습 진행 중

26 - 100 에폭부터: teacher forcing 0.5 -> 0.4, lr 5e-4 -> 3e-4, resize 2/3 없애고 풀사이즈, gaussian noise 없앴

27 - 136 에폭 이후로 val SA 뛰어서 teacher forcing 0.4 -> 0.3, lr 3e-4 -> 1e-4

28 -

29

30

31

32 - 김현우

33 - 가로 이미지 세로로 변형 : 0.051 상승

34 - ASTRN 튜닝하면서 실아

35

36

37 - 조호성

38 - CSTR 실패. 기존의 코드 적용한거는 22에폭까지 0나왔고, 논문 구현으로 코드 작성한거는 0.1을 넘지못함.

39

40

41 - 서준배

42 - adaptive 성능이 괜찮게 나옴

43 - hidden_dim, filter_dim 소폭 증가해서 진행중이

44 python

45 SATRN:

46 encoder:

47 hidden_dim: 300

48 filter_dim: 700 # 600

49 layer_num: 6

50 head num: 8

beam search를 적용할 수 있을까?

검색해보니 RNN 계열만 beam search를 쓰던데, CSTR은 CNN 계열이라 못쓸까??

다른 걸 적용할 수 있을까?

그나마 hyperparameter 조정??

아니면 단계별 학습??

윤준호

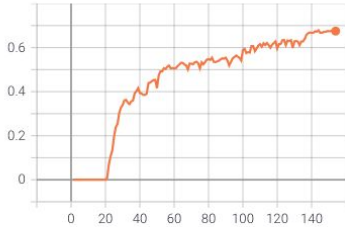
CNN depth 32, cosine annealing, augmentations 모델 조금씩 튜닝하면서 계속 학습 진행 중

100 에폭부터: teacher forcing 0.5 -> 0.4, lr 5e-4 -> 3e-4, resize 2/3 없애고 풀사이즈, gaussian noise 없앴

136 에폭 이후로 val SA 뛰어서 teacher forcing 0.4 -> 0.3, lr 3e-4 -> 1e-4

validation_sentence_accuracy

tag: validation_sentence_accuracy



김현우

가로 이미지 세로로 변형 : 0.051 상승

ASTRN 튜닝하면서 실아

조호성

CSTR 실패. 기존의 코드 적용한거는 22에폭까지 0나왔고, 논문 구현으로 코드 작성한거는 0.1을 넘지못함.

회의록

Aa 제목	월	화	수	목	금
1주차	5/24	5/25	5/26	5/27	5/28
2주차	5/31	6/1	6/2	6/3	6/4
3주차	6/7	6/8	6/9	6/10	6/11
4주차	6/14	6/15	-	-	-

boostcamp^{ai}tech Networking Day

© NAVER Connect Foundation

1. 문제 정의와 계획

실험 결과 공유 (구글 스프레드시트)

캠퍼	LB SA	LB WER	Val SA	Val WER	backbone	model	input_size	epochs	batch_size	cell-type	forcing-ratio	dropout	optimizer / lr	scheduler	Augmentation + 기타 설명
서준배	0.4483	0.1646	0.4526	0.1526	CNN	Attention	(128, 128)	50	96	LSTM	0.5	0.1	Adam	cycle (default)	
배철환	0.3398	0.309	0.3623	0.2448	CNN	SATRN		50	36	-	1	0.3	Adam		
임기홍	0.6396	0.1136	0.62193	0.09731	CNN	SATRN		50(30 제출)	36	-	0.5	0.1	Adam		가장 기본 (서버 폭파)
서준배	0.4723	0.1523	0.47284	0.14214	CNN	Attention		50+50	96	LSTM	0.5	0.1	Adam		
윤준호	0.6240	0.1061	0.6307	0.0978	CNN	SATRN		37	36	-	0.5	0.1	AdamW		
김현우	0.0023 증가		0.4399	0.1585	CNN	Attention		50	96	LSTM	0.5	0.1	Adam		num_layer 2 로 설정하면 오히려 성능이 떨어짐 (양상블 시도 오타로 실패 ..로컬에서 오타 수정후 성공)
배철환	0.4508	0.1651	0.4552	0.1554	CNN	Attention		50	96	GRU	0.5	0.1	Adam		
서준배			0.5872	0.1091	CNN	SATRN		50	36	-	0.5	0.1	Adam		A.Rotate(limit=90,border_mode=cv2.BORDER_WRAP,p=0.5) resize이후 추가
배철환					CNN	SATRN		100	36	-	0.5	0.1	Adamp		이미지 binarization, A.Compose([A.RandomContrast((0.3, 0.6), p = 1), A.CLAHE(clip_limit=(9,11), p = 1), A.RandomRotate90(p = 1), A.Resize(options.input_size.height, options.input_size.width, p = 1)]),
조호성	0.6082	0.0915	0.6144	0.0852	CNN	Attention		48/50	128	LSTM	0.5	0.1	Adam		deformable conv적용
윤준호			0.5454	0.1429	CNN	SATRN		50	36	-	0.5	0.4	AdamW / 3e-4	cosine annealing	Resize, Rotate(limit=30,border_mode=cv2.BORDER_WRAP,p=0.5), OneOf([CLAHE,IAASharpen,IAAEmboss,RandomContrast,RandomBrightness],p=0.3), A.GaussNoise(p=0.5)
조호성	0.6893	0.0769	0.6923	0.0698	CNN	SATRN		53	32		0.5	0.1	Adam		deformable conv적용
임기홍	0.6838	0.0776	0.6868	0.0716	CNN	SATRN		50	36		0.5	0.1	Adam		Basic

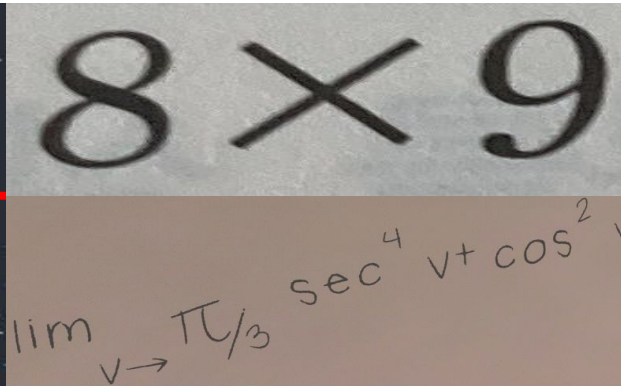
2. 실험 성과

- Dataset

외부 데이터 추가 ([Aida Calculus Math Handwriting Recognition Dataset](#))

- 토큰 규칙에 맞춰 변환
 - \lim_{\rightarrow} \rightarrow \lim 등
- 성능 향상
 - $0.7077 > 0.7124$

```
train_99995.jpg P \left( X = x_{\{ i \}} \right) = p_i
train_99996.jpg = 2 a^{\{ 2 \}} + 2 b^{\{ 2 \}} + 2 c^{\{ 2 \}} - 2 a b -
train_99997.jpg y - 3 =
train_99998.jpg \int_{\{ 0 \}}^{\{ k \}} \frac{e^{\{ x \}}}{\{ e^{\{ x \}} \}}
train_99999.jpg 5! = 120
bd85ee85-2549-4539-9cd9-122d0ea3dca2.jpg \lim_{\{ a \rightarrow \frac{\{ \}}{\{ \}} \}}
e97b0b1f-08bf-4c2b-86cb-f264af2160df.jpg \lim_{\{ w \rightarrow \pi / 5 \}}
3c72e1a1-c1de-4d67-a0ad-9f7afb6cee01.jpg e^{\lim_{\{ b \rightarrow 4 \}}
54459370-7cbe-423a-88bb-697d3864d12a.jpg \lim_{\{ \theta \rightarrow \pi \}}
da71ce9d-e1a1-4bb2-abce-77a29c45024f.jpg = \lim_{\{ x \rightarrow 8^{\{ \}} \}}
```



기본 Dataset (10만)

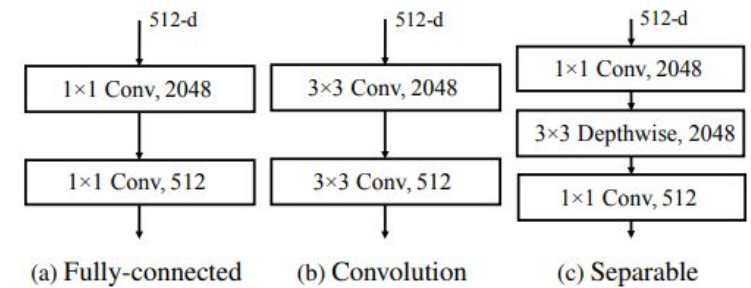
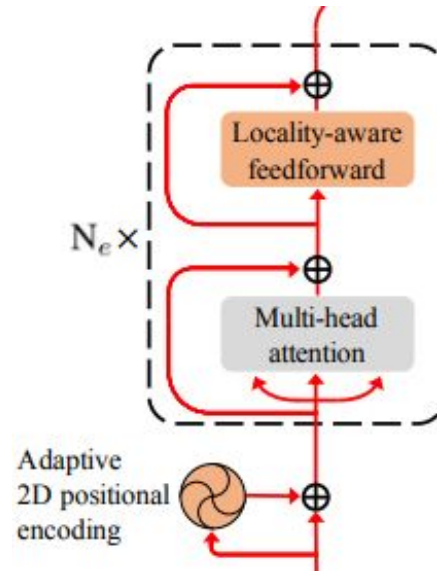
Aida Dataset (10만)

2. 실험 성과

- Model architecture

SATRN Model

- Adaptive 2D Positional Encoding
- Locality-Aware Feed Forward

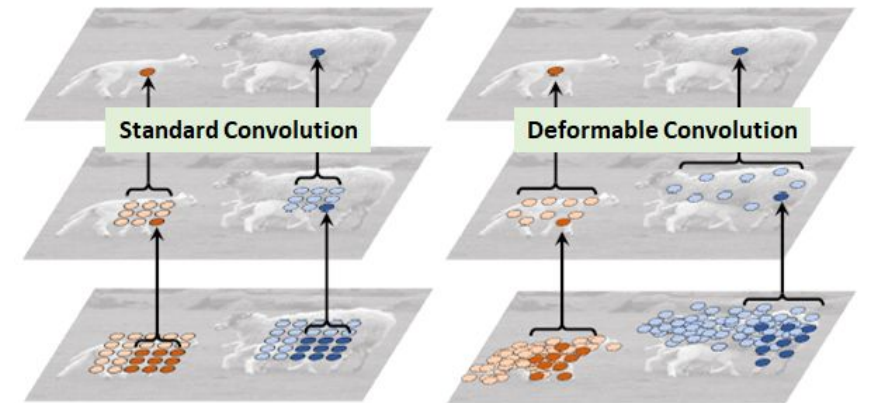
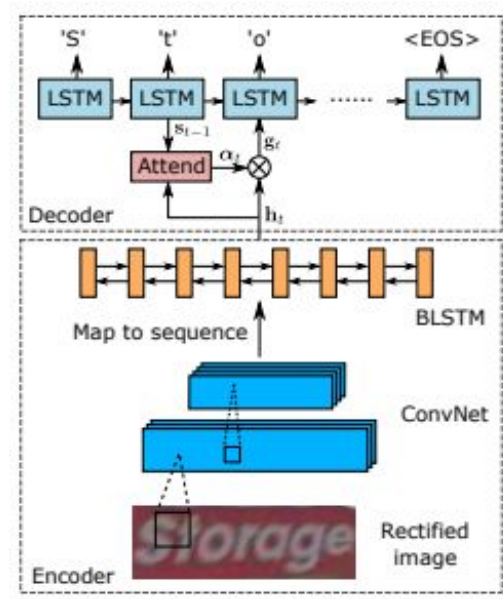


2. 실험 성과

- Model architecture

ASTER Model

- BI-LSTM
- Deformable Convolution



2. 실험 성과

- Hyperparameter Tuning & Ensemble

Hyperparameter Tuning

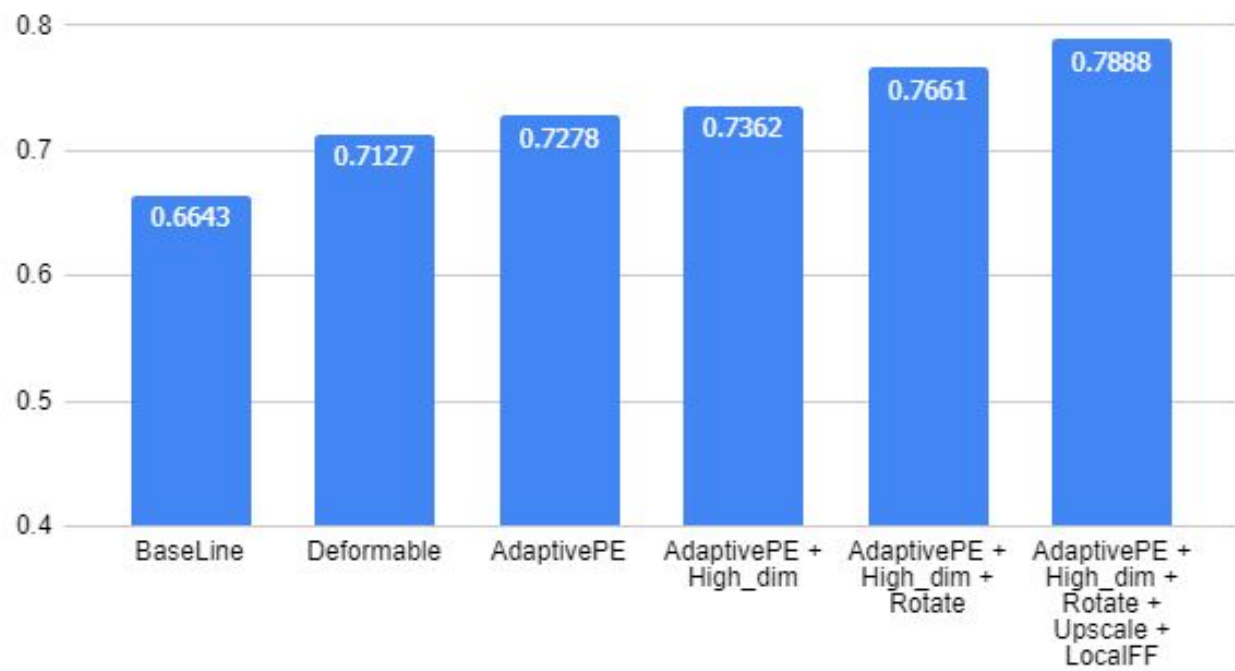
- Hidden & Filter Dimension 최적화
- CNN Depth 증가 (SATRN)
- Image Upscale
- Scheduler, Optimizer 탐색

Ensemble

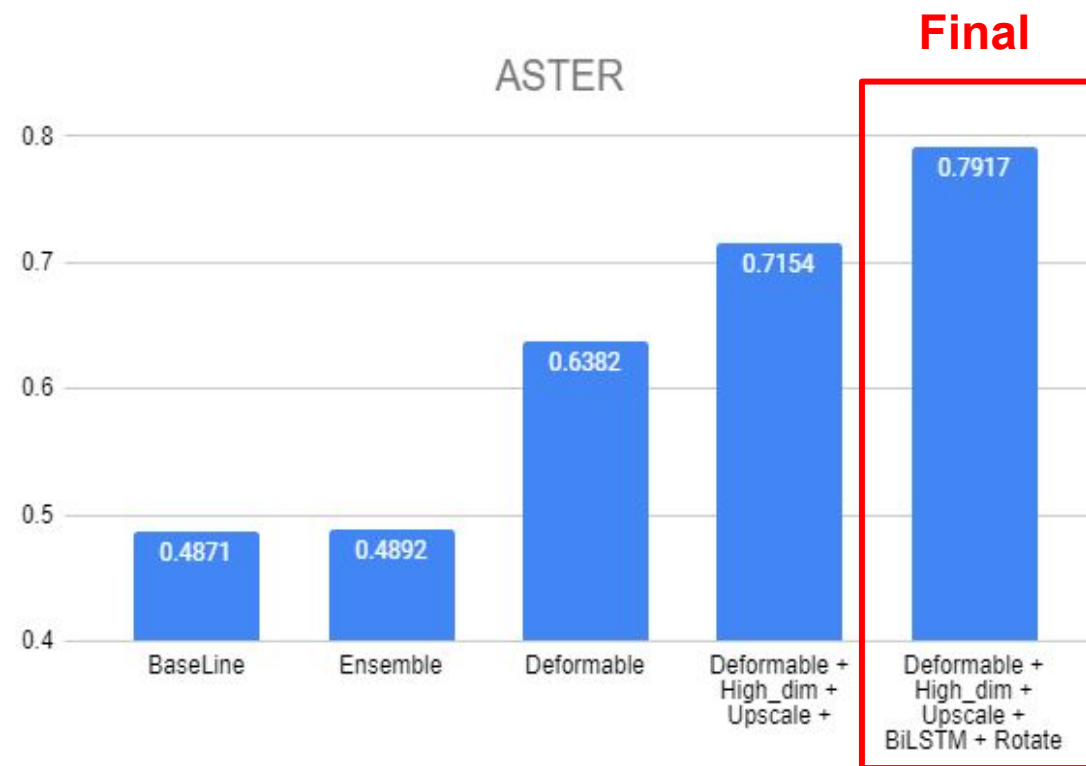
- SATRN(128,384) + SATRN(128,256) + Aster(80, 320)
 - Model-wise Multi-scale

2. 실험 성과

SATRN



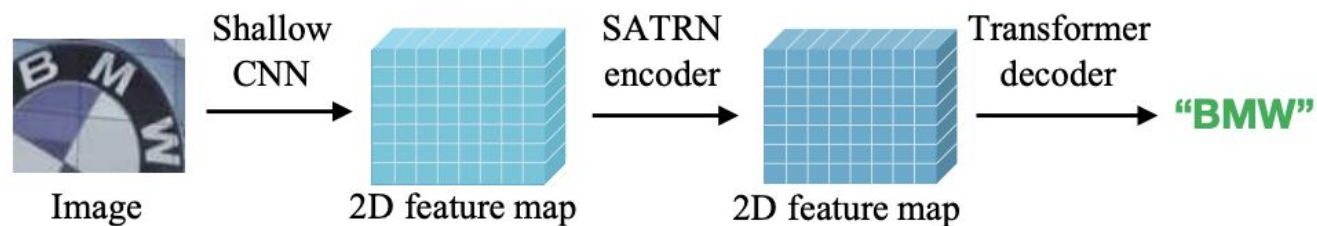
ASTER



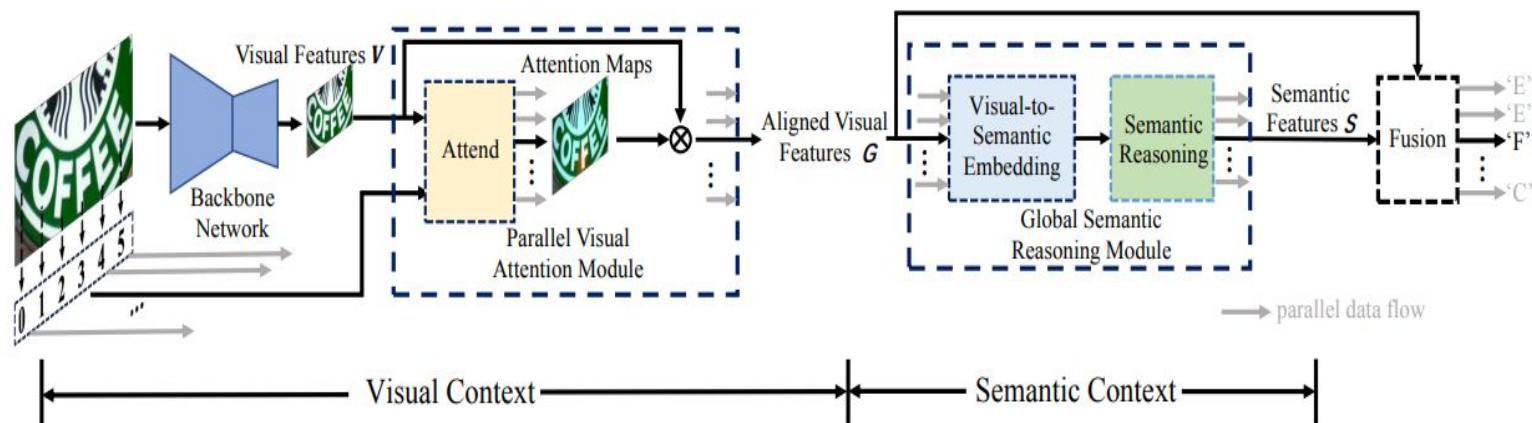
3. 모델의 탄생과 죽음

- SRN을 왜 선택했나

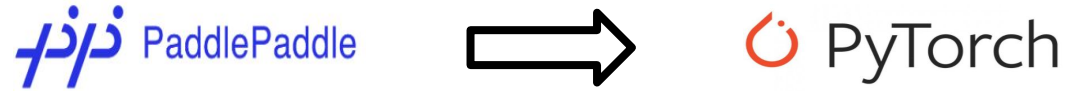
SATRN



SRN



3. 모델의 탄생과 죽음



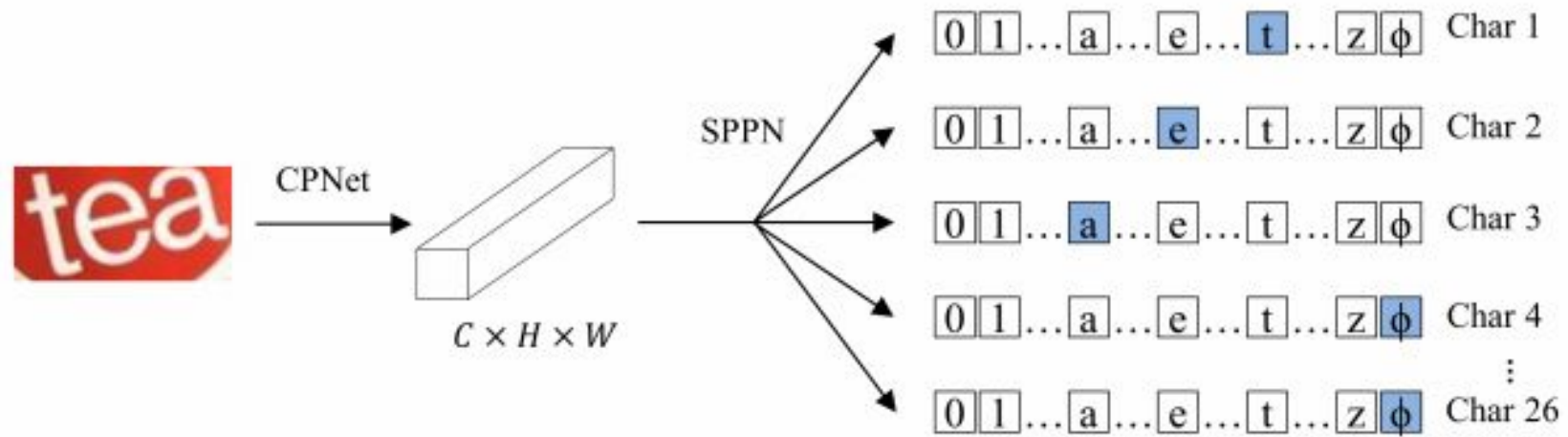
SRN 진행 과정

- 수렴하는 데 **2배 이상** 소요
 - LR, Scheduler 조정 후에도 시간은 개선되지 않음
- 실험 결과가 잘 적용되지 않아 **앙상블에서 제외**
 - Dim 증가 시 학습시간 증가

3. 모델의 탄생과 죽음

- CSTR

CNN으로만 이루어진 OCR Model



3. 모델의 탄생과 죽음

- 하지만 현실은...



3. 모델의 탄생과 죽음

- Underfitting 원인 분석

	Token size	Max Length
CSTR Paper	37	25
OURS	245	254

1 : 10 = Paper Params : Our Params

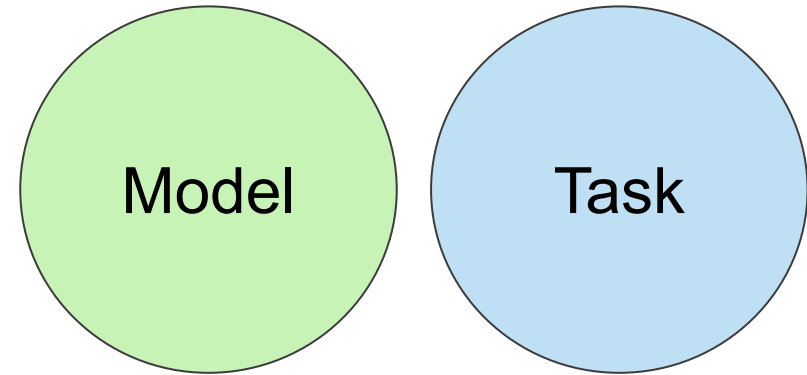
토큰 개수가 약 7배 차이!
문장최대길이가 약 10배 차이!

3. 모델의 탄생과 죽음

- 두 실패로부터 배운 것들

1. **모델 - 테스크 간 조합**을 철저하게 분석해야 함

2. **기간**을 잘 고려해서 계획을 세워야 함





show, Attend and Tell.pdf - Adobe Acrobat Reader DC (32-bit)

파일(F) 편집(E) 보기(V) 서명(S) 창(W) 도움말(H)

홈 도구 Show, Attend and T... x mai_03.pdf

128%

Neural Image Caption Generation with Visual Attention

weighted context into the system. The whole model is smooth and differentiable under the deterministic attention, so learning end-to-end is trivial by using standard back-propagation.

Learning the deterministic attention can also be understood as approximately optimizing the marginal likelihood in Equation 10 under the attention location random variable s_t from Sec. 4.1. The hidden activation of LSTM \mathbf{h}_t is a linear projection of the stochastic context vector $\hat{\mathbf{z}}_t$ followed by tanh non-linearity. To the first order Taylor approximation, the expected value $\mathbb{E}_{p(s_t|a)}[\mathbf{h}_t]$ is equal to computing \mathbf{h}_t using a single forward prop with the expected context vector $\mathbb{E}_{p(s_t|a)}[\hat{\mathbf{z}}_t]$. Considering Eq. 7, let $\mathbf{n}_t = \mathbf{L}_o(\mathbf{E}\mathbf{y}_{t-1} + \mathbf{L}_h\mathbf{h}_t + \mathbf{L}_z\hat{\mathbf{z}}_t)$. $\mathbf{n}_{t,i}$ denotes \mathbf{n}_t computed by setting the random variable $\hat{\mathbf{z}}$ value to \mathbf{a}_i . We define the normalized weighted geometric mean for the softmax k^{th} word prediction:

the scalar β .

Concretely, the model is trained end-to-end by minimizing the following penalized negative log-likelihood:

$$L_d = -\log \left(\frac{1}{C} \sum_i P(\mathbf{y}|\mathbf{x}) \right) + \lambda \sum_i^L \left(1 - \sum_t^C \alpha_{ti} \right)^2 \quad (14)$$

4.3. Training Procedure

Both variants of our attention model were trained with stochastic gradient descent using adaptive learning rate algorithms. For the Flickr8k dataset, we found that RMSPProp (Tieleman & Hinton, 2012) worked best, while for Flickr30k/MS COCO dataset we used the recently proposed Adam algorithm (Kingma & Ba, 2014).

To create the annotations \mathbf{a}_i used by our decoder, we used the Oxford VGGnet (Simonyan & Zisserman, 2014) pre-

'Bates' 검색

PDF 내보내기

PDF 편집

PDF 만들기

Adobe Acrobat Pro

유료 구독을 통해서 파일을 열고 다른 파일 형식으로 변환할 수 있습니다.

추가 정보

주석

파일 결합

PDF 압축

교정

Acrobat Pro D PDF 변환

무료 평가판

Q & A



End of Document

Thank You.

