Clova Developers

# Deep Learning Research of NAVER Clova for AI-Enhanced Business

2 July 2019

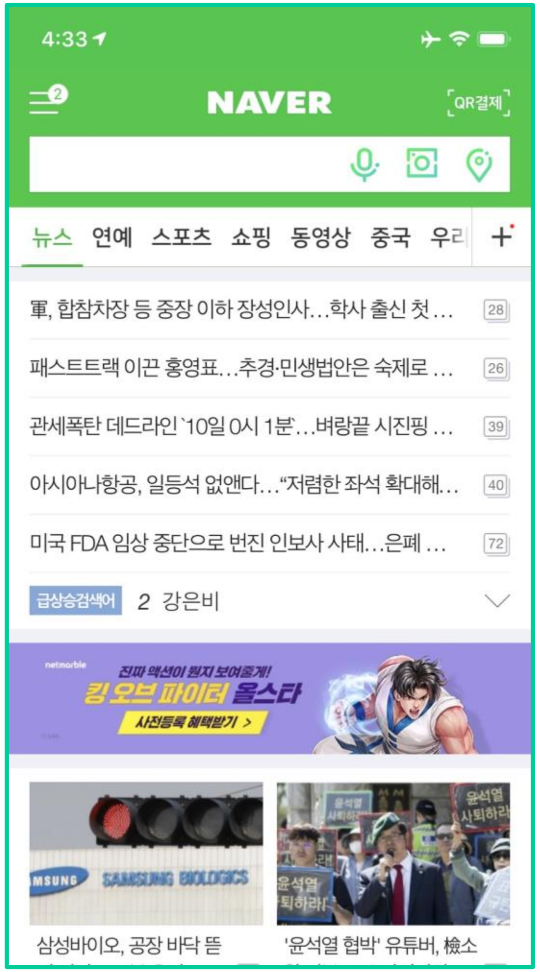
NVidia AI Conference

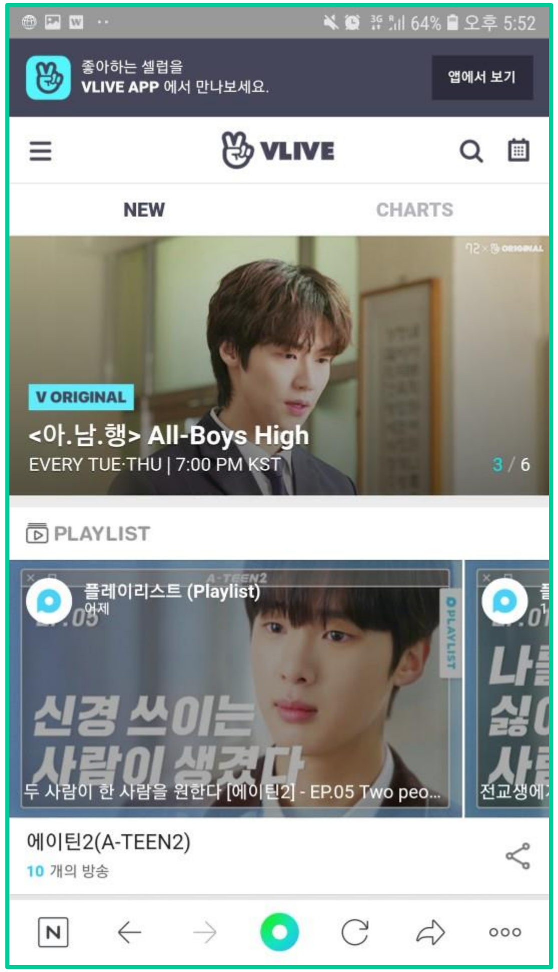
**Jung-Woo Ha, PhD**

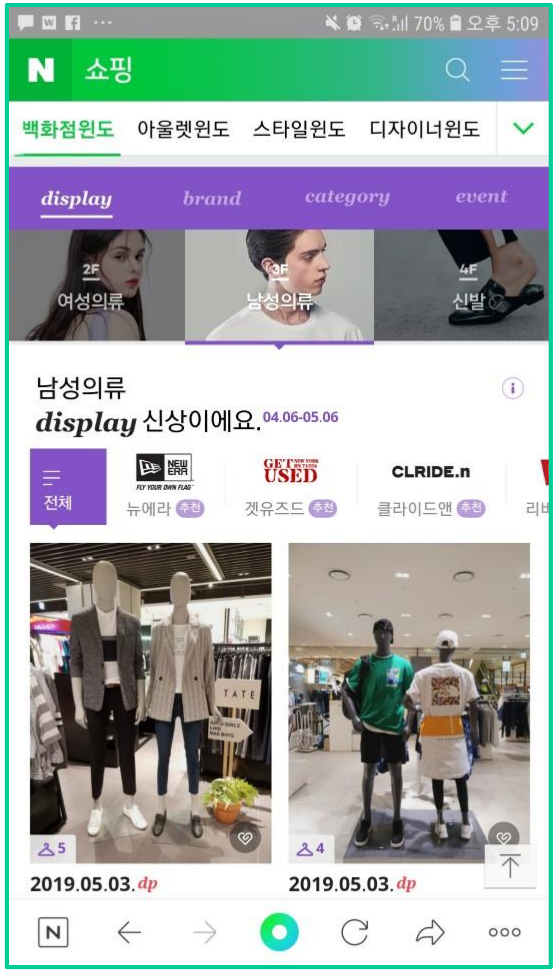
**Research Head, Clova AI, NAVER & LINE**

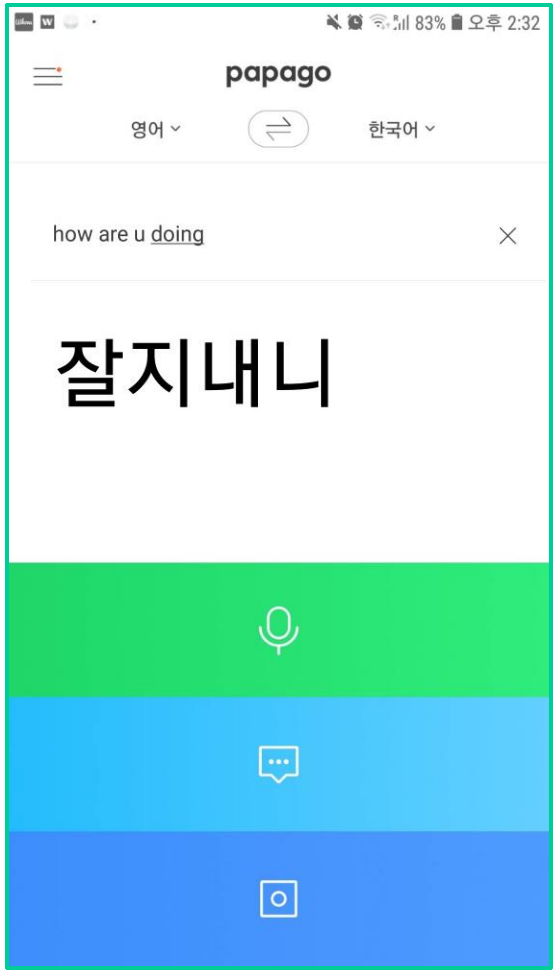
Clova

# NAVER


4:33
NAVER
QR결제
뉴스 연예 스포츠 쇼핑 동영상 중국 우리
軍, 합참차장 등 중장 이하 장성인사…학사 출신 첫…
28
패스트트랙 이끈 홍영표…추경·민생법안은 숙제로…
26
관세폭탄 데드라인 '10일 0시 1분'…벼랑끝 시진핑…
39
아시아나항공, 일등석 없앤다…"저렴한 좌석 확대해…
40
미국 FDA 임상 중단으로 번진 인보사 사태…은폐…
72
급상승검색어
2 강은비
진짜 액션이 뭔지 보여줄게!
킹 오브 파이터 올스타
사전등록 혜택받기 >
SAMSUNG BIOLOGICS
삼성바이오, 공장 바닥 뜯
'윤석열 협박' 유튜버, 檢소

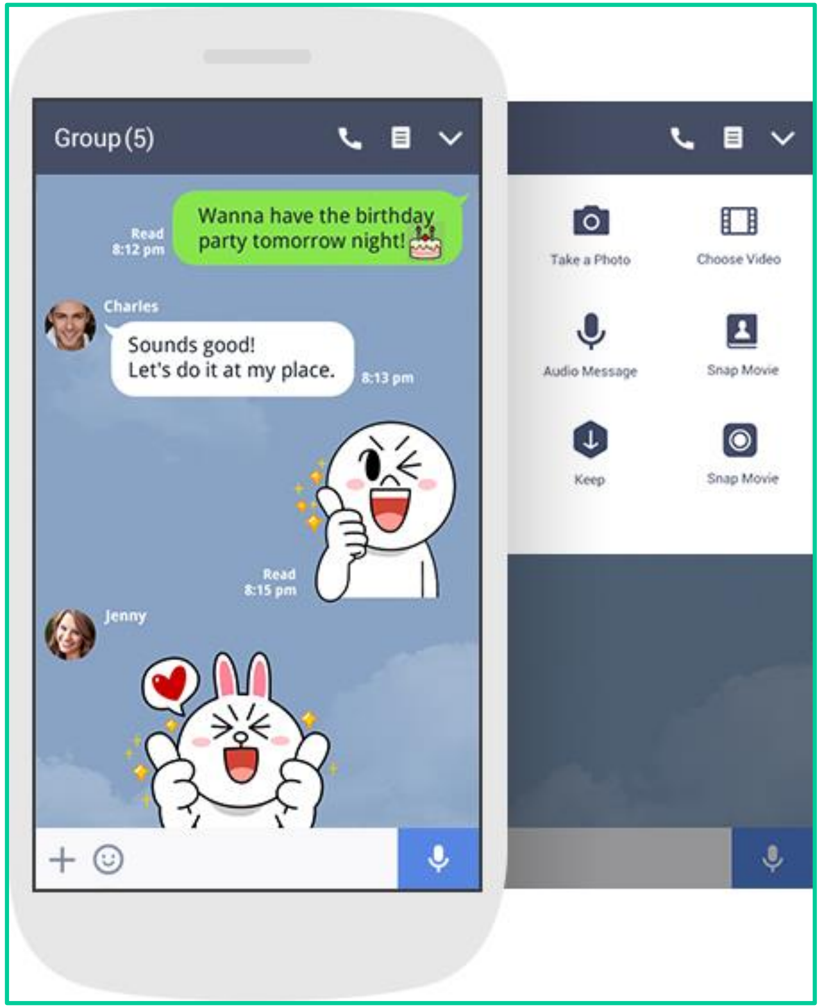

오후 5:52
좋아하는 셀럽을
VLIVE APP 에서 만나보세요.
앱에서 보기
VLIVE
NEW
CHARTS
V ORIGINAL
<아.남.행> All-Boys High
EVERY TUE·THU | 7:00 PM KST
3 / 6
PLAYLIST
플레이리스트 (Playlist)
어제
신경 쓰이는
사람이 생겼다
두 사람이 한 사람을 원한다 [에이틴2] - EP.05 Two peo…
에이틴2(A-TEEN2)
10 개의 방송


오후 5:09
N 쇼핑
백화점윈도 아울렛윈도 스타일윈도 디자이너윈도
display
brand
category
event
2F
여성의류
3F
남성의류
4F
신발
남성의류
display 신상이에요. 04.06-05.06
전체
뉴에라 추천
겟유즈드 추천
CLRIDE.n
클라이드앤 추천
2019.05.03. dp
2019.05.03. dp


오후 2:32
papago
영어
한국어
how are u doing
잘지내니

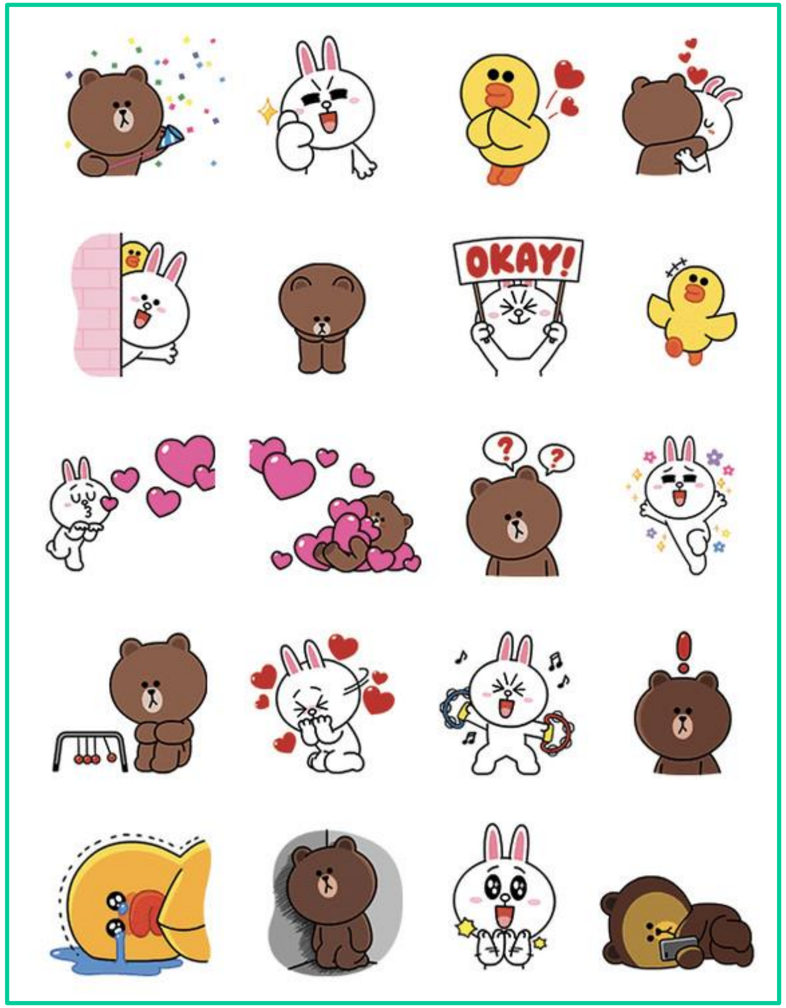
Clova

# LINE



Group (5)
Wanna have the birthday party tomorrow night!
Read 8:12 pm
Charles
Sounds good! Let's do it at my place.
8:13 pm
Read 8:15 pm
Jenny
Take a Photo
Choose Video
Audio Message
Snap Movie
Keep
Snap Movie


7:30 PM
LINE NEWS
総合
エンタメ
トレンド
スポーツ
時事
銀座に格安オーダーシューズ
コンフェデ杯開会式でサンバ
さしこ、センター曲決定

OKAY!

# NAVER & LINE

## NAVER

Korea No.1 Internet Platform Company

## Clova

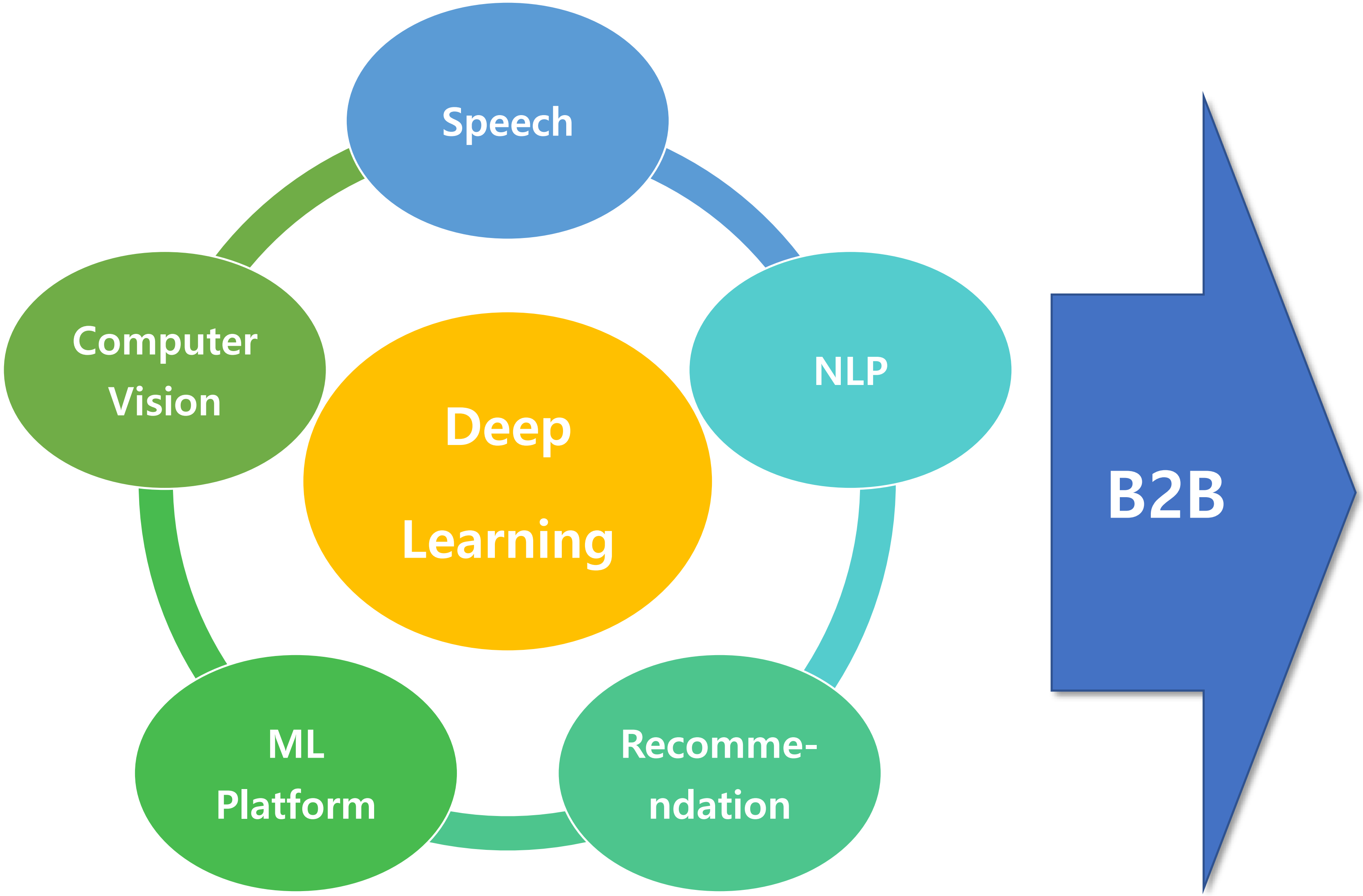
## LINE

Global Messenger Platform Company Based on Japan

# Vision of Clova: AI for Everyone

Innovative AI Technologies → Practical & Valuable AI Services → B2B / B2C → Better World

# Clova AI Core Technologies to B2B




- OCR
- FaceSDK
- Chatbot
- AI ARS
- Speech Recognition
- Voice Synthesis
- NSML

# AI Technology Hierarchy

**Lightweight** | **AutoML**

Application Logics + API

Task-specific Models

(Fine-tuning, Transfer learning, Distillation, Domain adaptation, …)

Powerful Pretrained Models

(Regularizer, DataAug, LR scheduling, Curriculum learning, …)

# Pretrained Models

# Distillation: Student Better Than A Teacher

[Heo et al. Arxiv 2019]

**[CIFAR-100]**

| Setup | Compression type | Teacher network | Student network | # of params teacher | # of params student | Compress ratio |
|---|---|---|---|---|---|---|
| (a) | Depth | WideResNet 28-4 | WideResNet 16-4 | 5.87M | 2.77M | 47.2% |
| (b) | Channel | WideResNet 28-4 | WideResNet 28-2 | 5.87M | 1.47M | 25.0% |
| (c) | Depth & channel | WideResNet 28-4 | WideResNet 16-2 | 5.87M | 0.70M | 11.9% |
| (d) | Different architecture | WideResNet 28-4 | ResNet 56 | 5.87M | 0.86M | 14.7% |
| (e) | Different architecture | PyramidNet-200 (240) | WideResNet 28-4 | 26.84M | 5.87M | 21.9% |
| (f) | Different architecture | PyramidNet-200 (240) | PyramidNet-110 (84) | 26.84M | 3.91M | 14.6% |

| Setup | Teacher | Baseline | KD [8] | FitNets [22] | AT [30] | Jacobian [26] | FT [14] | AB [7] | Proposed |
|---|---|---|---|---|---|---|---|---|---|
| (a) | 21.09 | 22.72 | 21.69 | 21.85 | 22.07 | 22.18 | 21.72 | 21.36 | **20.89** |
| (b) | 21.09 | 24.88 | 23.43 | 23.94 | 23.80 | 23.70 | 23.41 | 23.19 | **21.98** |
| (c) | 21.09 | 27.32 | 26.47 | 26.30 | 26.56 | 26.71 | 25.91 | 26.02 | **24.08** |
| (d) | 21.09 | 27.68 | 26.76 | 26.35 | 26.66 | 26.60 | 26.20 | 26.04 | **24.44** |
| (e) | 15.57 | 21.09 | 20.97 | 22.16 | 19.28 | 20.59 | 19.04 | 20.46 | **17.80** |
| (f) | 15.57 | 22.58 | 21.68 | 23.79 | 19.93 | 23.49 | 19.53 | 20.89 | **18.89** |

# Distillation: Student Better Than A Teacher

[Heo et al. Arxiv 2019]

## [ImageNet-1k]

| Network | # of param (ratio) | Method | Top-1 error(%) | Top-5 error(%) |
|---|---|---|---|---|
| ResNet152 | 60.19M | Teacher | 21.69 | 5.95 |
| ResNet50 | 25.56M (42.5%) | Baseline | 23.72 | 6.97 |
| | | AT [30] | 22.75 | 6.35 |
| | | FT [14] | 22.80 | 6.49 |
| | | AB [7] | 23.47 | 6.94 |
| | | Proposed | **21.65** | **5.83** |
| ResNet50 | 25.56M | Teacher | 23.84 | 7.14 |
| MobileNet | 4.23M (16.5%) | Baseline | 31.13 | 11.24 |
| | | AT [30] | 30.44 | 10.67 |
| | | FT [14] | 30.12 | 10.50 |
| | | AB [7] | 31.11 | 11.29 |
| | | Proposed | **28.75** | **9.66** |

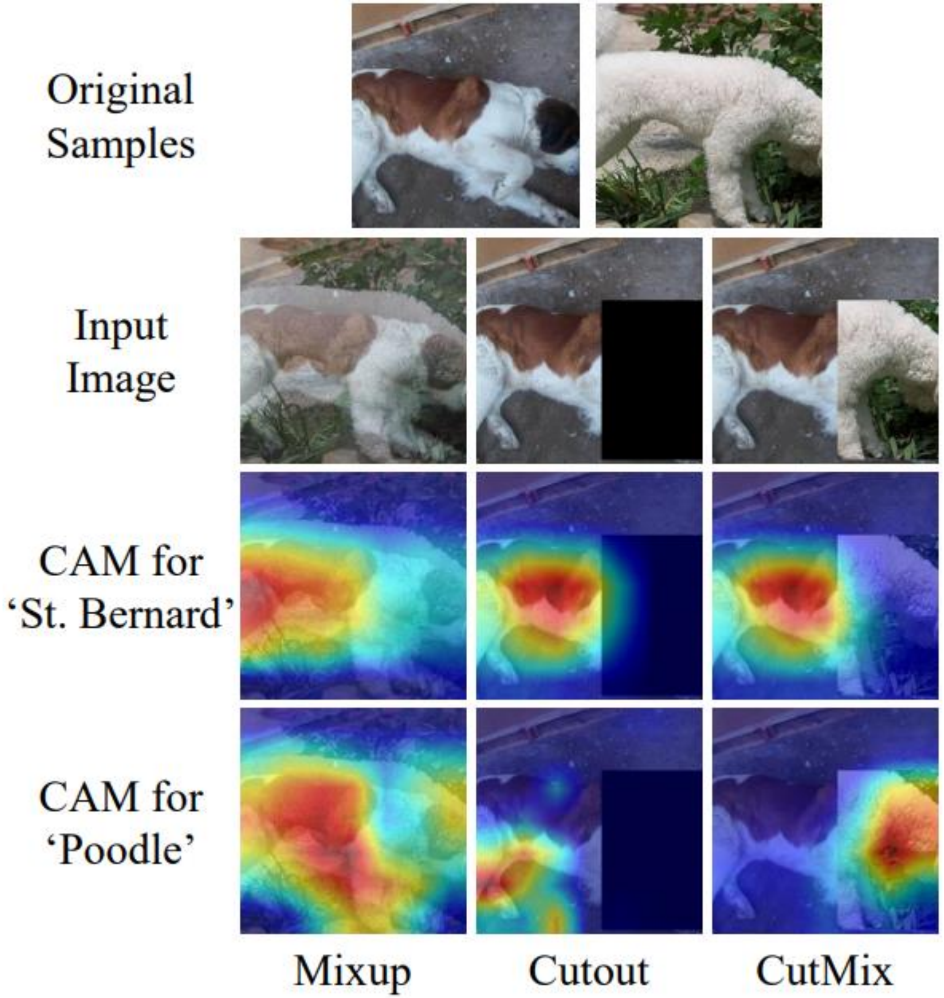
## [Other Tasks: Object Detection & Segmentation]

| Network | # of params | Method | mAP(%) |
|---|---|---|---|
| ResNet50-SSD | 36.7M | Teacher (T1) | 76.79 |
| VGG-SSD | 26.3M | Teacher (T2) | 77.50 |
| ResNet18-SSD | 20.0M | Baseline | 71.61 |
| | | Proposed-T1 | **73.08** |
| | | Proposed-T2 | 72.38 |
| MobileNet -SSD lite | 6.5M | Baseline | 67.58 |
| | | Proposed-T1 | **68.54** |
| | | Proposed-T2 | 68.45 |

| Backbone | # of params | Method | mIoU |
|---|---|---|---|
| ResNet101 | 59.3M | Teacher | 77.39 |
| ResNet18 | 16.6M (28.0%) | Baseline | 71.79 |
| | | Proposed | **73.24** |
| MobileNetV2 | 5.8M (9.8%) | Baseline | 68.44 |
| | | Proposed | **71.36** |

# CutMix: New Robust Data Augmentation

[Yun et al. Arxiv 2019]




| | ResNet-50 | Mixup [46] | Cutout [2] | CutMix |
|---|---|---|---|---|
| Image |  | | | |
| Label | Dog 1.0 | Dog 0.5<br>Cat 0.5 | Dog 1.0 | Dog 0.6<br>Cat 0.4 |
| ImageNet Cls (%) | 76.3 (+0.0) | 77.4 (+1.1) | 77.1 (+0.8) | **78.4 (+2.1)** |
| ImageNet Loc (%) | 46.3 (+0.0) | 45.8 (-0.5) | 46.7 (+0.4) | **47.3 (+1.0)** |
| Pascal VOC Det (mAP) | 75.6 (+0.0) | 73.9 (-1.7) | 75.1 (-0.5) | **76.7 (+1.1)** |

# CutMix: New Robust Data Augmentation

Clova

[Yun et al. Arxiv 2019]

| Model | # Params | Top-1 Err (%) | Top-5 Err (%) |
|---|---|---|---|
| ResNet-152* | 60.3 M | 21.69 | 5.94 |
| ResNet-101 + SE Layer* [14] | 49.4 M | 20.94 | 5.50 |
| ResNet-101 + GE Layer* [13] | 58.4 M | 20.74 | 5.29 |
| ResNet-50 + SE Layer* [14] | 28.1 M | 22.12 | 5.99 |
| ResNet-50 + GE Layer* [13] | 33.7 M | 21.88 | 5.80 |
| ResNet-50 (Baseline) | 25.6 M | 23.68 | 7.05 |
| ResNet-50 + Cutout [2] | 25.6 M | 22.93 | 6.66 |
| ResNet-50 + StochDepth [16] | 25.6 M | 22.46 | 6.27 |
| ResNet-50 + Mixup [46] | 25.6 M | 22.58 | 6.40 |
| ResNet-50 + Manifold Mixup [40] | 25.6 M | 22.50 | 6.21 |
| ResNet-50 + DropBlock* [7] | 25.6 M | 21.87 | 5.98 |
| ResNet-50 + Feature CutMix | 25.6 M | 21.80 | 6.06 |
| ResNet-50 + CutMix | 25.6 M | **21.60** | **5.90** |

| | Baseline | Mixup | Cutout | CutMix |
|---|---|---|---|---|
| Top-1 Acc (%) | 8.2 | 24.4 | 11.5 | **31.0** |

Table 11: Top-1 accuracy after FGSM white-box attack on ImageNet validation set.

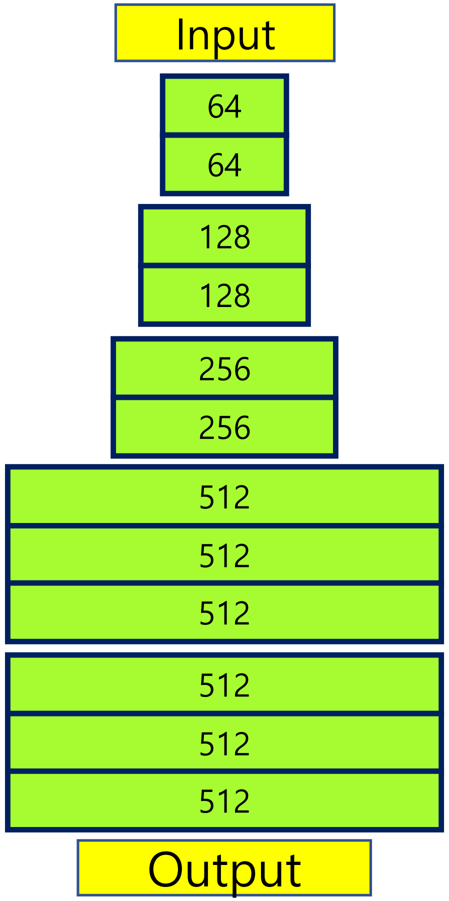
| Method | TNR at TPR 95% | AUROC | Detection Acc. |
|---|---|---|---|
| Baseline | 26.3 (+0) | 87.3 (+0) | 82.0 (+0) |
| Mixup | 11.8 (-14.5) | 49.3 (-38.0) | 60.9 (-21.0) |
| Cutout | 18.8 (-7.5) | 68.7 (-18.6) | 71.3 (-10.7) |
| CutMix | **69.0** (+42.7) | **94.4** (+7.1) | **89.1** (+7.1) |

| Backbone Network | ImageNet Cls Top-1 Error (%) | Detection | | Image Captioning | |
|---|---|---|---|---|---|
| | | SSD [23] (mAP) | Faster-RCNN [29] (mAP) | NIC [41] (BLEU-1) | NIC [41] (BLEU-4) |
| ResNet-50 (Baseline) | 23.68 | 76.7 (+0.0) | 75.6 (+0.0) | 61.4 (+0.0) | 22.9 (+0.0) |
| Mixup-trained | 22.58 | 76.6 (-0.1) | 73.9 (-1.7) | 61.6 (+0.2) | 23.2 (+0.3) |
| Cutout-trained | 22.93 | 76.8 (+0.1) | 75.0 (-0.6) | 63.0 (+1.6) | 24.0 (+1.1) |
| CutMix-trained | 21.60 | **77.6** (+0.9) | **76.7** (+1.1) | **64.2** (+2.8) | **24.9** (+2.0) |

# Lightweight CNN Architecture Design

- Here, the problem is **how to design feature-map sizes when the # of parameters is limited**.

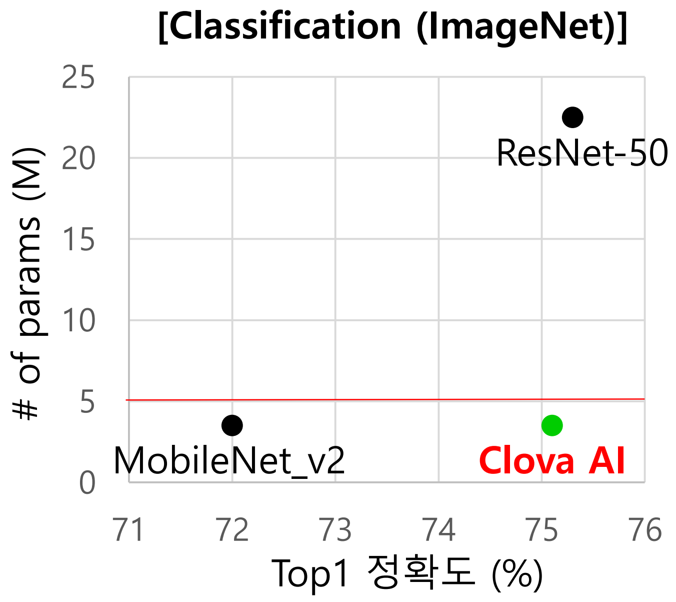
- The performance could be improved by weight layer reallocation.

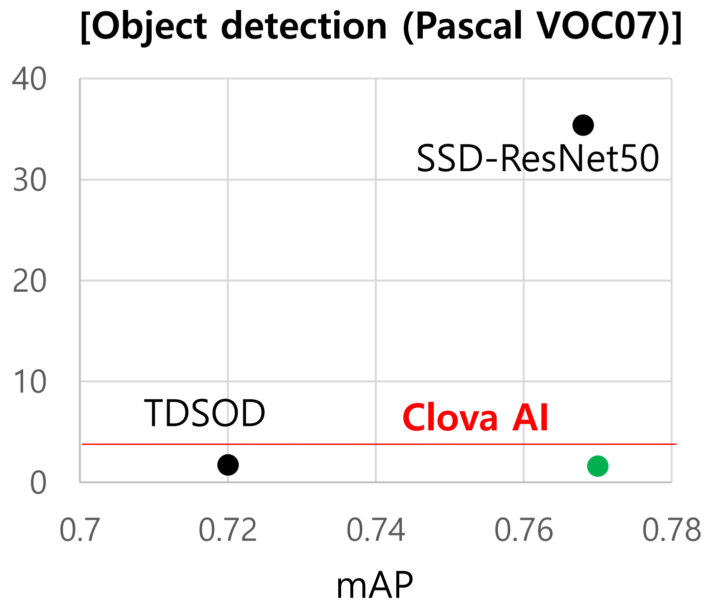


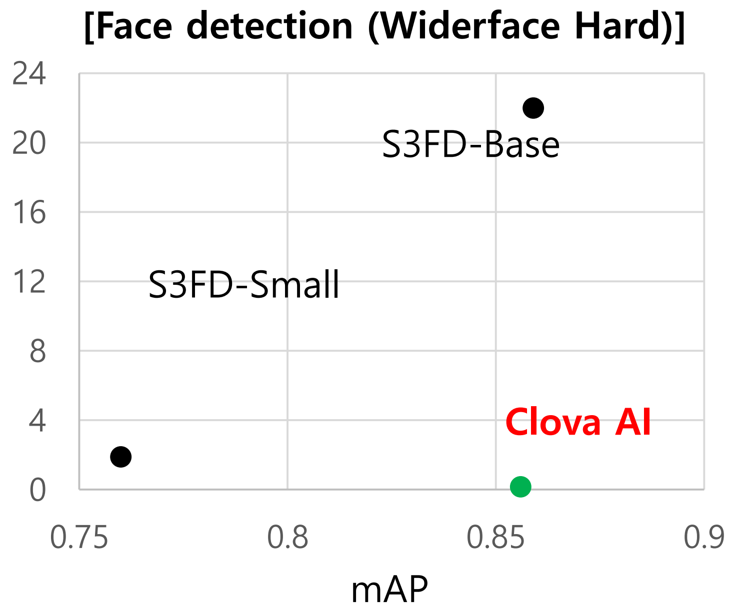
Clova

# SOTA Lightweight Image Models

[Han et al. 2019; Yun et al. 2019; Yoo et al. 2019]




NAVER | LINE

# Lightweight CNN Architecture Design

- Transfer to object detection task (finetuning)
  - Pascal VOC 07 test results (trained on VOC 0712 trainval):

| | Clova AI 경량화 | SSD - MobileNet_v2 |
|---|---|---|
| PASCAL: VOC 정확도(모델크기) | 77.0 (5.4M) | 70.1 (5.4M) |

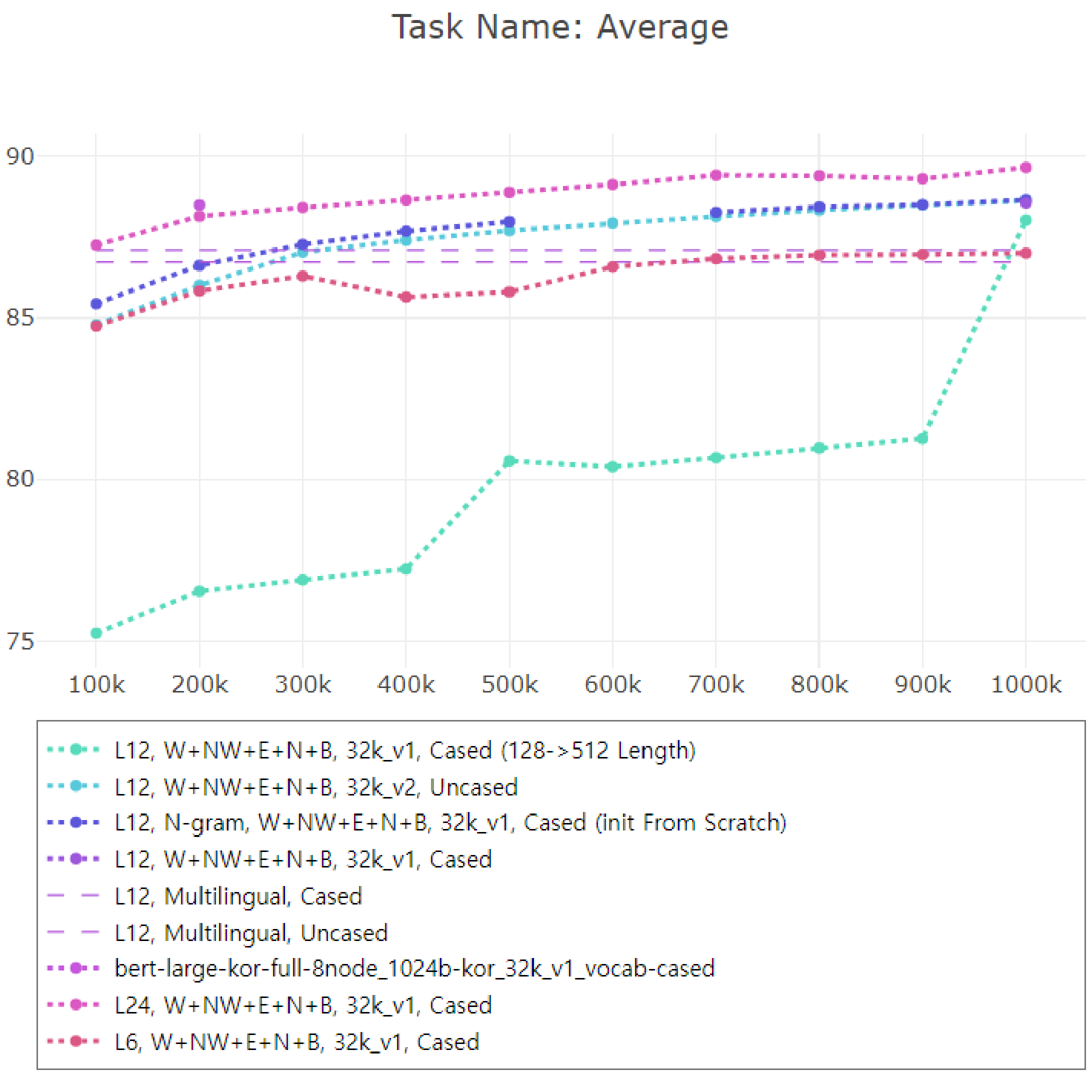
- Transfer to text detection task (Lite-CRAFT)
  - ICDAR-13 test results:

| Backbone | # of params | Hmean(%) |
|---|---|---|
| VGG-16 BN | **20.8 M** | **91.5** |
| Ours | **2.3 M** | **91.0** |
| Ours | **2.1 M** | **89.0** |

# LaRva: Language Representation by Clova

Clova

- BigLM based on BERT
- New task definition
  - GLUE vs. KLUE and JLUE
- New encoding, corpus-level curriculum learning, n-gram maksing
- Distributed learning for LaRva
- Fast LaRva



# Task-specific Models

Clova

# Speech Enhancement

[Orignal]


[Enhanced]

Clova

# Audio-Visual Speech Enhancement

- Speech separation given lip regions in the video

**[Afouras et al. Interspeech 2018]**


# HDTS: SOTA Speech Synthesis

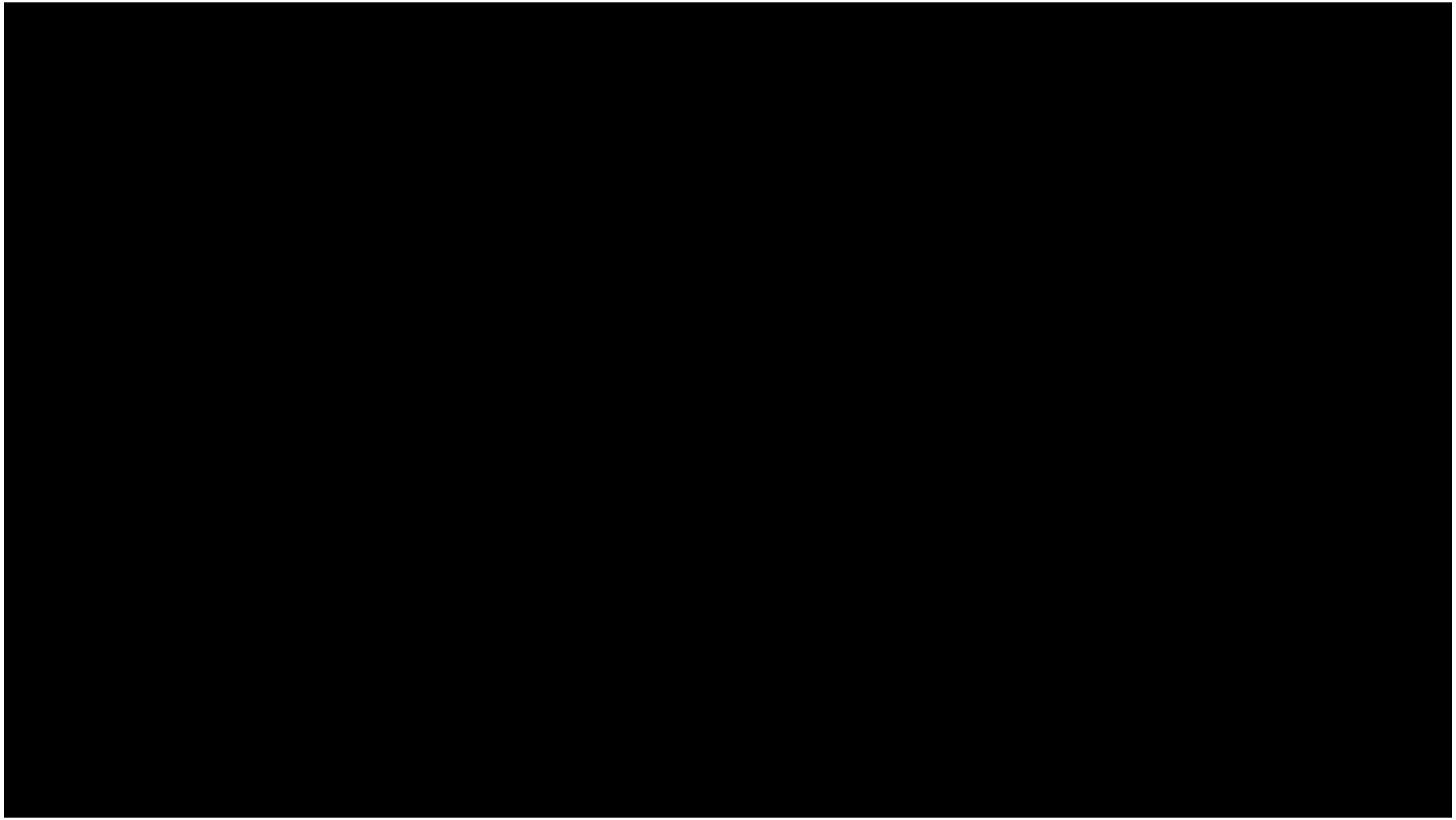
Required recorded voice data: 10hrs → 4hrs → 40mins

# Detection in OCR



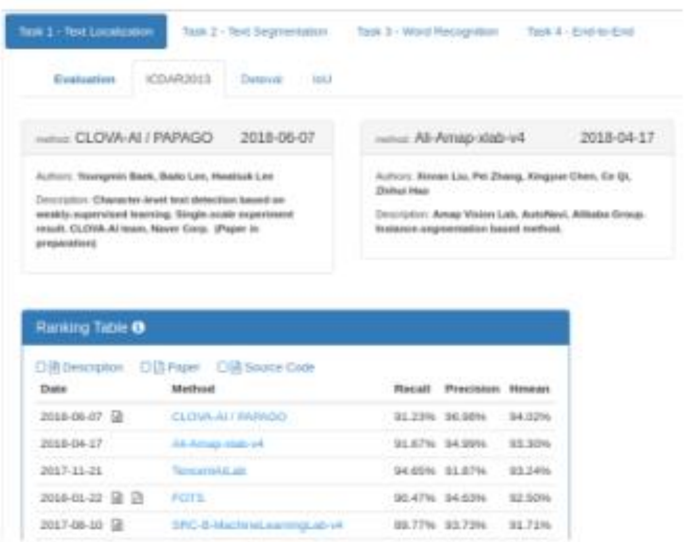
[Baek et al. CVPR 2019]

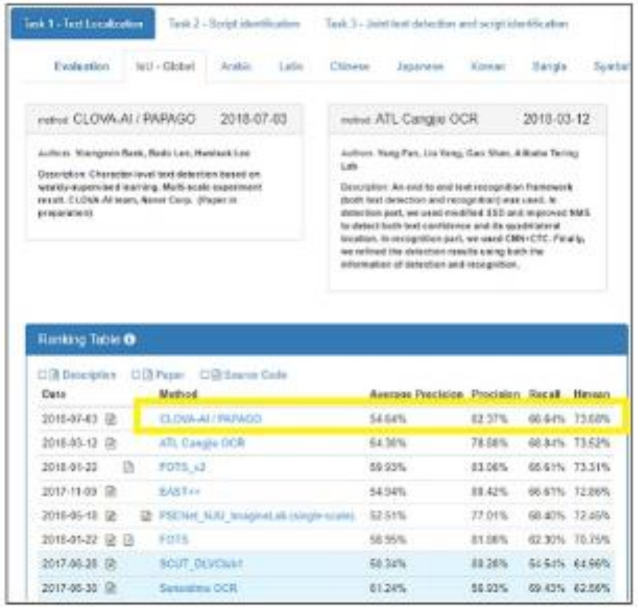
CRAFT: Character region awareness for text detection

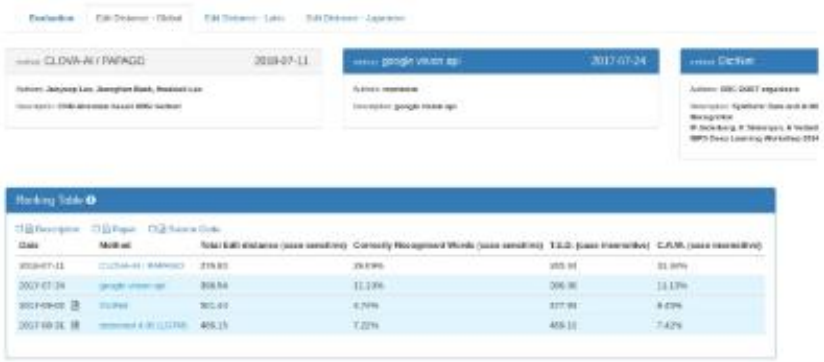


# OCR Challenges

- The 1st rank in 4 Leaderboards @ ICDAR Challenges (Jan 2019)



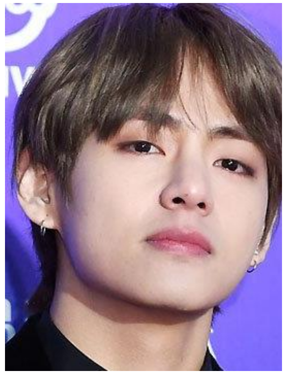




Clova

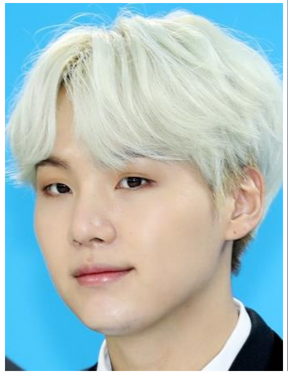
# AutoCut (BTS)

Source video

V



Suga



Jungkook



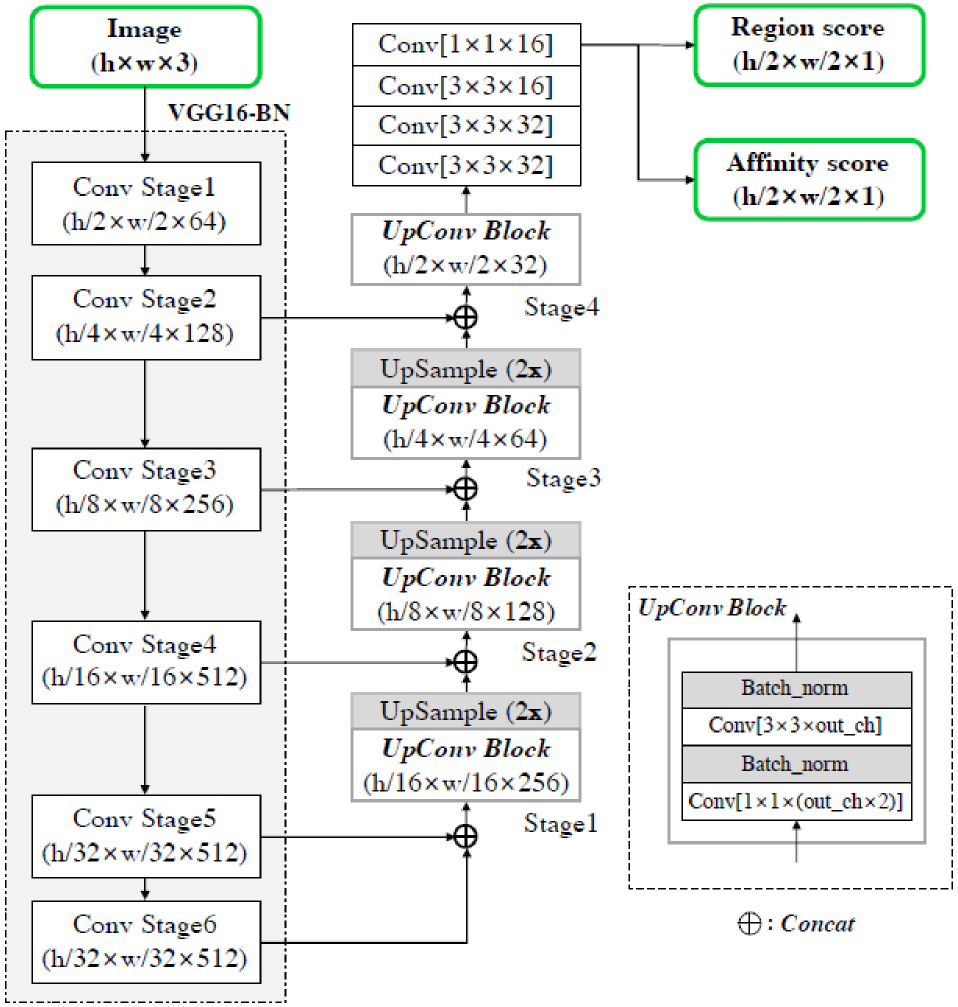
https://channels.vlive.tv/EF0205/home

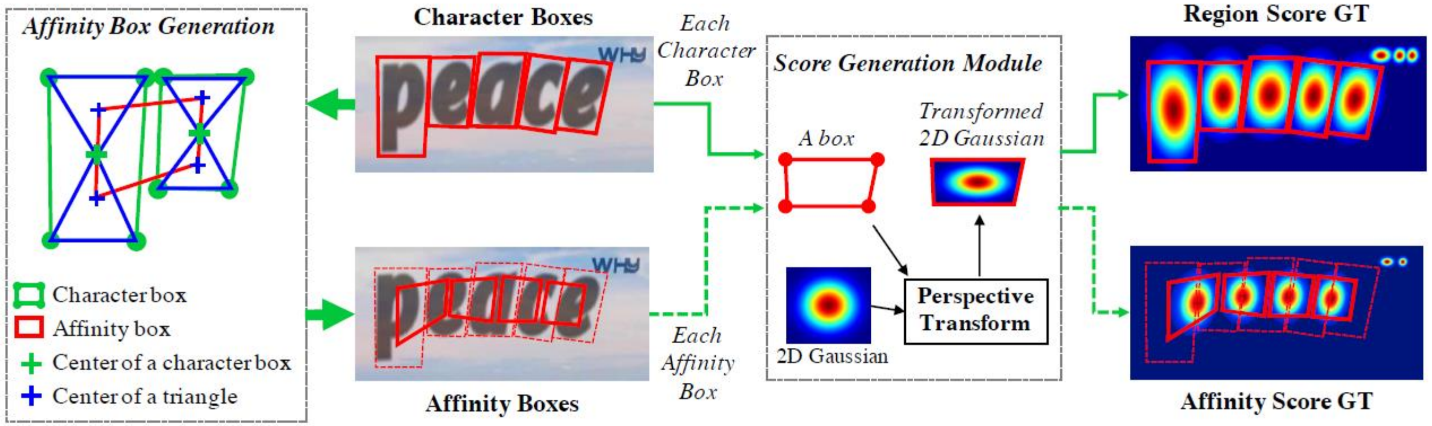
# AutoCAM (Black Pink)

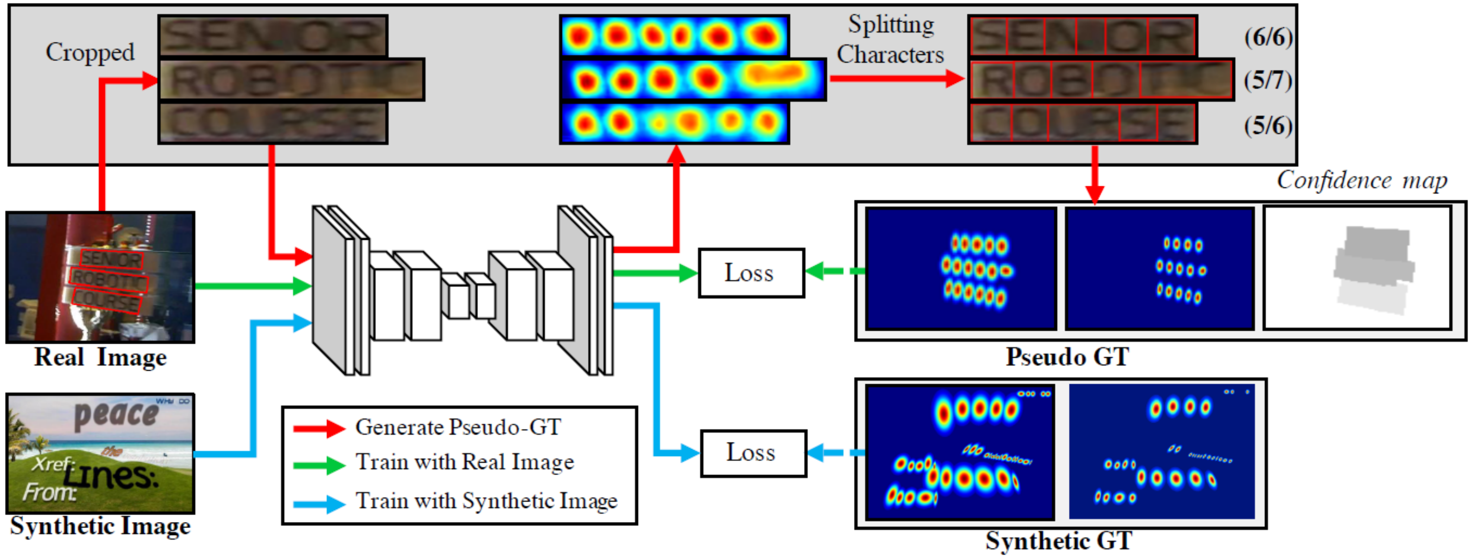
Applications of tracking, pose estimation, and person re-id in in-the-wild videos

# Detection in OCR

CRAFT: Character region awareness for text detection **[Baek et al. CVPR 2019]**



# Detection in OCR

[Baek et al. CVPR 2019]

CRAFT: Character region awareness for text detection


| Method | IC13(DetEval) | | | IC15 | | | IC17 | | | MSRA-TD500 | | | FPS |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | R | P | H | R | P | H | R | P | H | R | P | H | |
| Zhang et al. [39] | 78 | 88 | 83 | 43 | 71 | 54 | - | - | - | 67 | 83 | 74 | 0.48 |
| Yao et al. [37] | 80.2 | 88.8 | 84.3 | 58.7 | 72.3 | 64.8 | - | - | - | 75.3 | 76.5 | 75.9 | 1.61 |
| SegLink [32] | 83.0 | 87.7 | 85.3 | 76.8 | 73.1 | 75.0 | - | - | - | 70 | 86 | 77 | 20.6 |
| SSTD [8] | 86 | 89 | 88 | 73 | 80 | 77 | - | - | - | - | - | - | 7.7 |
| Wordsup [12] | 87.5 | 93.3 | 90.3 | 77.0 | 79.3 | 78.2 | - | - | - | - | - | - | 1.9 |
| EAST* [40] | - | - | - | 78.3 | 83.3 | 80.7 | - | - | - | 67.4 | 87.3 | 76.1 | 13.2 |
| He et al. [11] | 81 | 92 | 86 | 80 | 82 | 81 | - | - | - | 70 | 77 | 74 | 1.1 |
| R2CNN [13] | 82.6 | 93.6 | 87.7 | 79.7 | 85.6 | 82.5 | - | - | - | - | - | - | 0.4 |
| TextSnake [24] | - | - | - | 80.4 | 84.9 | 82.6 | - | - | - | 73.9 | 83.2 | 78.3 | 1.1 |
| TextBoxes++* [17] | 86 | 92 | 89 | 78.5 | 87.8 | 82.9 | - | - | - | - | - | - | 2.3 |
| *EAA* [10] | *87* | *88* | *88* | *83* | *84* | *83* | - | - | - | - | - | - | - |
| *Mask TextSpotter* [25] | *88.1* | *94.1* | *91.0* | *81.2* | *85.8* | *83.4* | - | - | - | - | - | - | 4.8 |
| PixelLink* [4] | 87.5 | 88.6 | 88.1 | 82.0 | 85.5 | 83.7 | - | - | - | 73.2 | 83.0 | 77.8 | 3.0 |
| RRD* [19] | 86 | 92 | 89 | 80.0 | 88.0 | 83.8 | - | - | - | 73 | 87 | 79 | 10 |
| Lyu et al.* [26] | 84.4 | 92.0 | 88.0 | 79.7 | 89.5 | 84.3 | **70.6** | 74.3 | 72.4 | 76.2 | 87.6 | 81.5 | 5.7 |
| *FOTS* [21] | - | - | *87.3* | *82.0* | *88.8* | *85.3* | *57.5* | *79.5* | *66.7* | - | - | - | 23.9 |
| **CRAFT(ours)** | **93.1** | **97.4** | **95.2** | **84.3** | **89.8** | **86.9** | 68.2 | **80.6** | **73.9** | **78.2** | **88.2** | **82.9** | 8.6 |

# Recognition in OCR

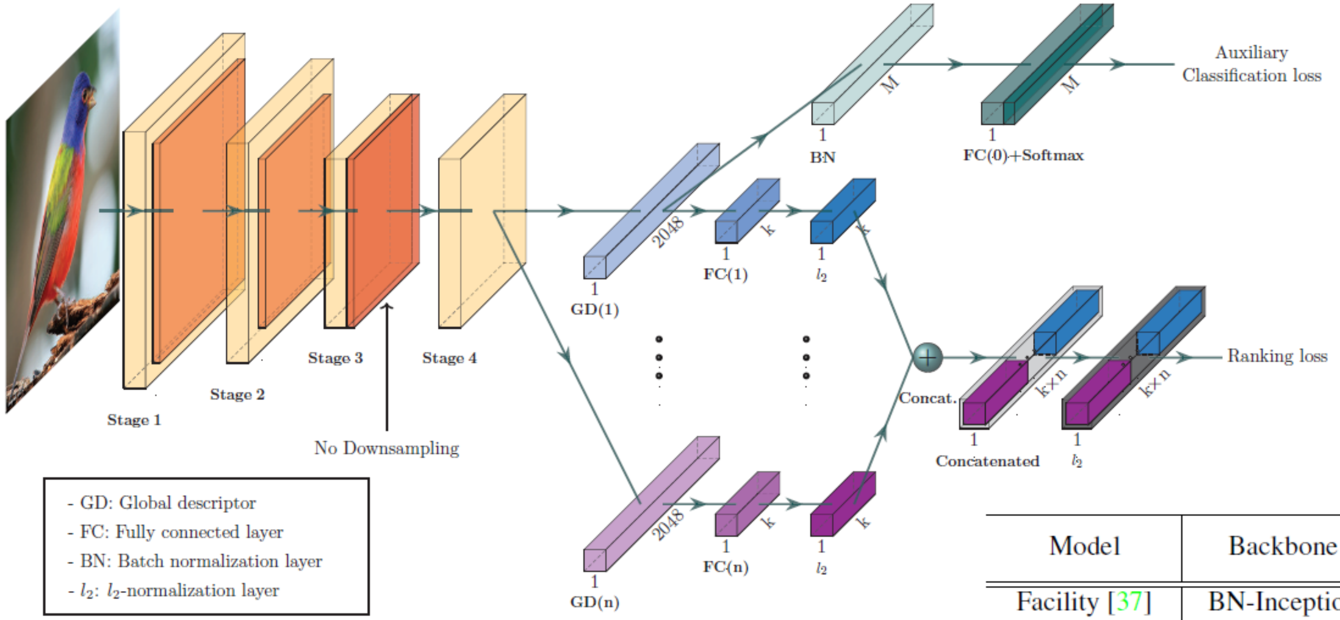
- Full integration for practical OCR

**[Baek et al. ArXiv 2019]**

| | Model | Year | Train data | IIIT | SVT | IC03 | | IC13 | | IC15 | | SP | CT | Time | params |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | 3000 | 647 | 860 | 867 | 857 | 1015 | 1811 | 2077 | 645 | 288 | ms/image | $\times 10^6$ |
| Reported results | CRNN [23] | 2015 | MJ | 78.2 | 80.8 | 89.4 | – | – | 86.7 | – | – | – | – | 160 | 8.3 |
| | RARE [24] | 2016 | MJ | 81.9 | 81.9 | 90.1 | – | 88.6 | – | – | – | 71.8 | 59.2 | $<2$ | – |
| | R2AM [15] | 2016 | MJ | 78.4 | 80.7 | 88.7 | – | – | 90.0 | – | – | – | – | 2.2 | – |
| | STAR-Net [17] | 2016 | MJ+PRI | 83.3 | 83.6 | 89.9 | – | – | 89.1 | – | – | 73.5 | – | – | – |
| | GRCNN [26] | 2017 | MJ | 80.8 | 81.5 | 91.2 | – | – | – | – | – | – | – | – | – |
| | ATR [28] | 2017 | PRI+C | – | – | – | – | – | – | – | – | **75.8** | 69.3 | – | – |
| | FAN [4] | 2017 | MJ+ST+C | 87.4 | 85.9 | – | 94.2 | – | 93.3 | 70.6 | – | – | – | – | – |
| | Char-Net [16] | 2018 | MJ | 83.6 | 84.4 | **91.5** | – | 90.8 | – | – | 60.0 | 73.5 | – | – | – |
| | AON [5] | 2018 | MJ+ST | 87.0 | 82.8 | – | 91.5 | – | – | – | **68.2** | 73.0 | **76.8** | – | – |
| | EP [2] | 2018 | MJ+ST | 88.3 | **87.5** | – | 94.6 | – | **94.4** | **73.9** | – | – | – | – | – |
| | Rosetta [3] | 2018 | PRI | – | – | – | – | – | – | – | – | – | – | – | – |
| | SSFL [18] | 2018 | MJ | **89.4** | 87.1 | – | **94.7** | **94.0** | – | – | – | 73.9 | 62.5 | – | – |
| Our experiment | CRNN [23] | 2015 | MJ+ST | 82.9 | 81.6 | 93.1 | 92.6 | 91.1 | 89.2 | 69.4 | 64.2 | 70.0 | 65.5 | 4.4 | 8.3 |
| | RARE [24] | 2016 | MJ+ST | 86.2 | 85.8 | 93.9 | 93.7 | 92.6 | 91.1 | 74.5 | 68.9 | 76.2 | 70.4 | 23.6 | 10.8 |
| | R2AM [15] | 2016 | MJ+ST | 83.4 | 82.4 | 92.2 | 92.0 | 90.2 | 88.1 | 68.9 | 63.6 | 72.1 | 64.9 | 24.1 | 2.9 |
| | STAR-Net [17] | 2016 | MJ+ST | 87.0 | 86.9 | 94.4 | 94.0 | 92.8 | 91.5 | 76.1 | 70.3 | 77.5 | 71.7 | 10.9 | 48.7 |
| | GRCNN [26] | 2017 | MJ+ST | 84.2 | 83.7 | 93.5 | 93.0 | 90.9 | 88.8 | 71.4 | 65.8 | 73.6 | 68.1 | 10.7 | 4.6 |
| | Rosetta [3] | 2018 | MJ+ST | 84.3 | 84.7 | 93.4 | 92.9 | 90.9 | 89.0 | 71.2 | 66.0 | 73.8 | 69.2 | 4.7 | 44.3 |
| | Our best model | | MJ+ST | **87.9** | **87.5** | **94.9** | **94.4** | **93.6** | **92.3** | **77.6** | **71.8** | **79.2** | **74.0** | 27.6 | 49.6 |

# SOTA Fashion Retrieval

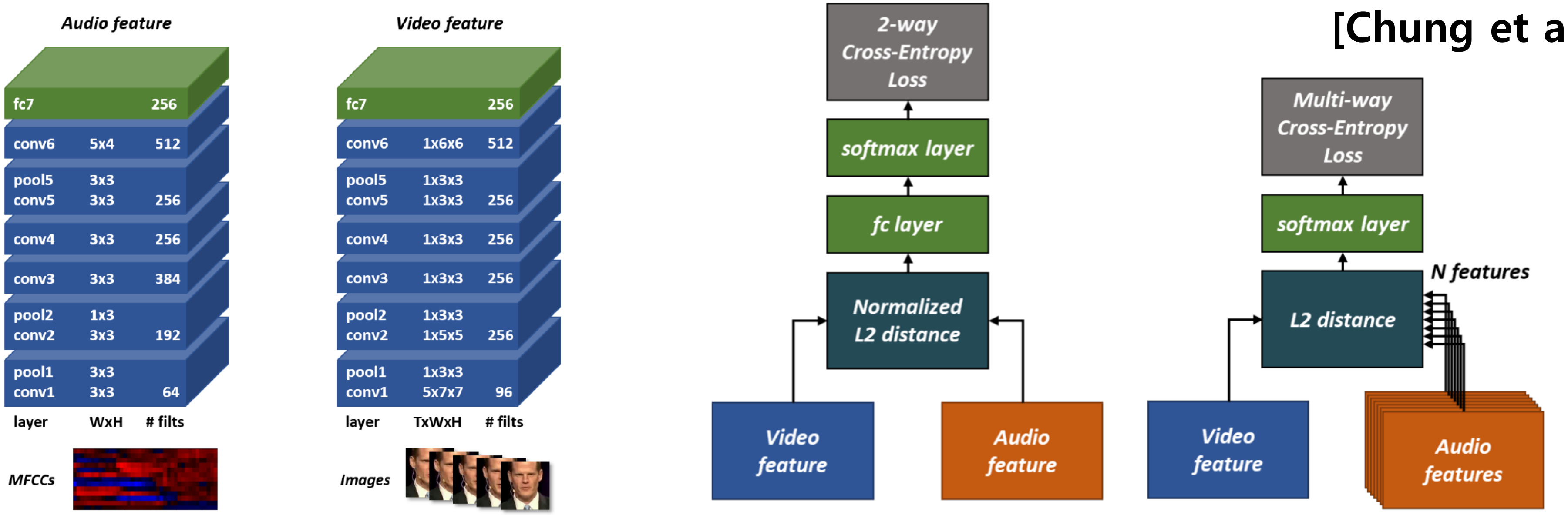
[Jeon et al. Arxiv 2019]




| Model | Backbone | Dim | SOP | | | | In-shop | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 1 | 10 | 100 | 1000 | 1 | 10 | 20 | 30 | 40 | 50 |
| Facility [37] | BN-Inception | 64 | 67.0 | 83.7 | 93.2 | - | - | - | - | - | - | |
| HTL [9] | BN-Inception | 512 | 74.8 | 88.3 | 94.8 | 98.4 | - | - | - | - | - | - |
| HTL [9] | BN-Inception | 128 | - | - | - | - | 80.9 | 94.3 | 95.8 | 97.2 | 97.4 | 97.8 |
| Margin [54] | ResNet-50 | 128 | 72.7 | 86.2 | 93.8 | 98.0 | - | - | - | - | - | - |
| ABE-8 [22] | GoogleNet‡ | 512 | 76.3 | 88.4 | 94.8 | 98.2 | 87.3 | 96.7 | 97.9 | 98.2 | 98.5 | 98.7 |
| BFE† [7] | ResNet-50‡ | 1536 | 83.0 | 93.3 | 97.3 | **99.2** | 89.1 | 96.3 | 97.6 | 98.5 | **99.1** | - |
| CGD (SG/GS) | BN-Inception | 64 | **75.6** | **89.0** | **95.5** | **98.6** | 86.6 | 96.3 | 97.4 | 97.9 | 98.2 | 98.4 |
| CGD (SG/ - ) | BN-Inception | 512 | **80.5** | **92.1** | **96.7** | **98.9** | - | - | - | - | - | - |
| CGD ( - /GS) | BN-Inception | 128 | - | - | - | - | **88.5** | **97.1** | **98.0** | **98.5** | **98.8** | **98.9** |
| CGD (SG/GS) | ResNet-50 | 128 | **81.0** | **92.2** | **96.8** | **99.1** | 88.4 | 97.2 | 98.1 | 98.4 | 98.7 | 98.8 |
| CGD (SG/GS) | ResNet-50‡ | 1536 | **83.9** | **93.8** | **97.5** | **99.2** | **90.9** | **98.0** | **98.7** | **99.0** | **99.1** | **99.2** |
| CGD (SG/GS) | ShuffleNet-v2 | 1536 | 78.7 | 90.9 | 96.1 | 98.8 | 86.1 | 96.9 | 97.8 | 98.4 | 98.6 | 98.7 |
| CGD (SG/GS) | SE-ResNet-50‡ | 1536 | **84.2** | **93.9** | 97.4 | **99.2** | **91.9** | **98.1** | **98.7** | **99.0** | **99.1** | **99.3** |

# Audio-Visual Speech Enhancement

[Chung et al. ICASSP 2019]




(a) Audio stream (b) Visual stream (b) AVE-Net (c) Proposed model

**Table 1.** Synchronization accuracy. **# Frames**: the number of visual frames for which the distances are averaged over.

| **# Frames** | SyncNet | AVE-Net | **Proposed** |
|---|---|---|---|
| 5 | 75.8% | 74.1% | **89.5%** |
| 7 | 82.3% | 80.4% | **92.1%** |
| 9 | 87.6% | 86.1% | **94.7%** |
| 11 | 91.8% | 90.6% | **96.1%** |
| 13 | 94.5% | 93.7% | **97.5%** |
| 15 | 96.1% | 95.5% | **98.1%** |

**Table 2.** Word accuracy of lip reading using various architectures and training methods.

| **Architecture** | **Method** | **Top-1 (%)** |
|---|---|---|
| MT-5 [15] | E2E | 66.8 |
| LF-5 [15] | E2E | 66.0 |
| LSTM-5 [15] | E2E | 65.4 |
| TC-5 | E2E | 71.5 |
| TC-5 | PT - SyncNet | 67.8 |
| TC-5 | PT - AVE-Net | 66.7 |
| TC-5 | PT - **Proposed** | **71.6** |

Clova

# Super-Real Style Transfer

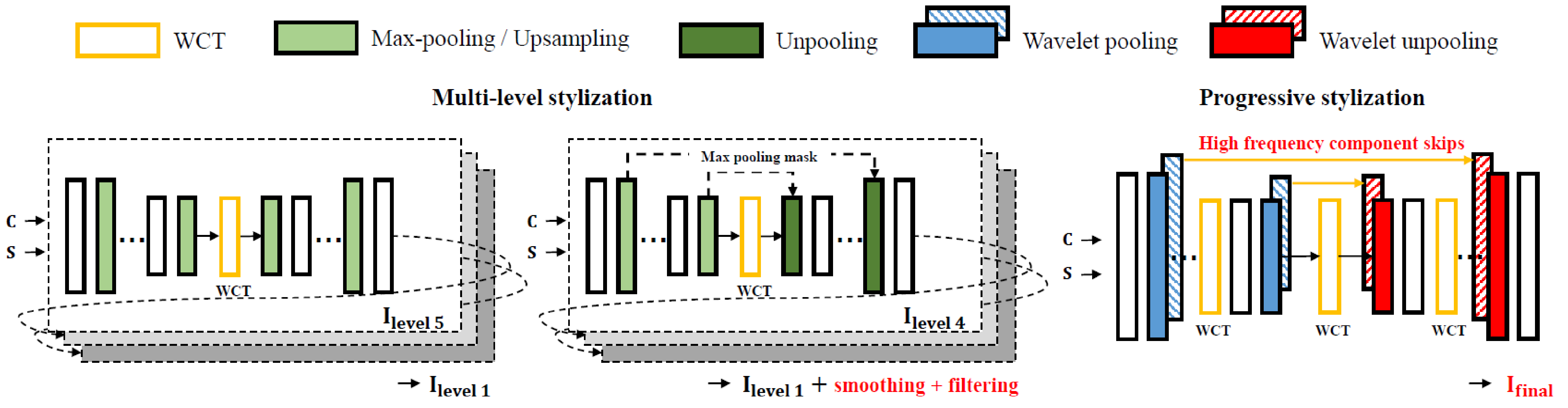
[Yoo et al. Arxiv 2019]


[WCT] [PhotoWCT wo pp] [WCT2(ours)]

# Super-Real Style Transfer

- Wavelet Pooling for Perfect Reconstruction

[Yoo et al. Arxiv 2019]




https://github.com/clovaai/wct2

# EXTD: EXtremely Tiny face Detector





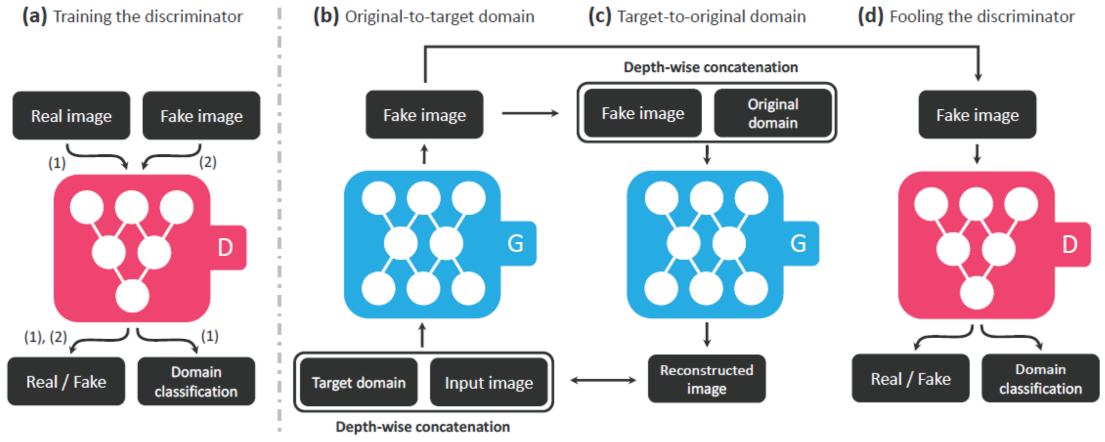


[EXTD: 0.16M]

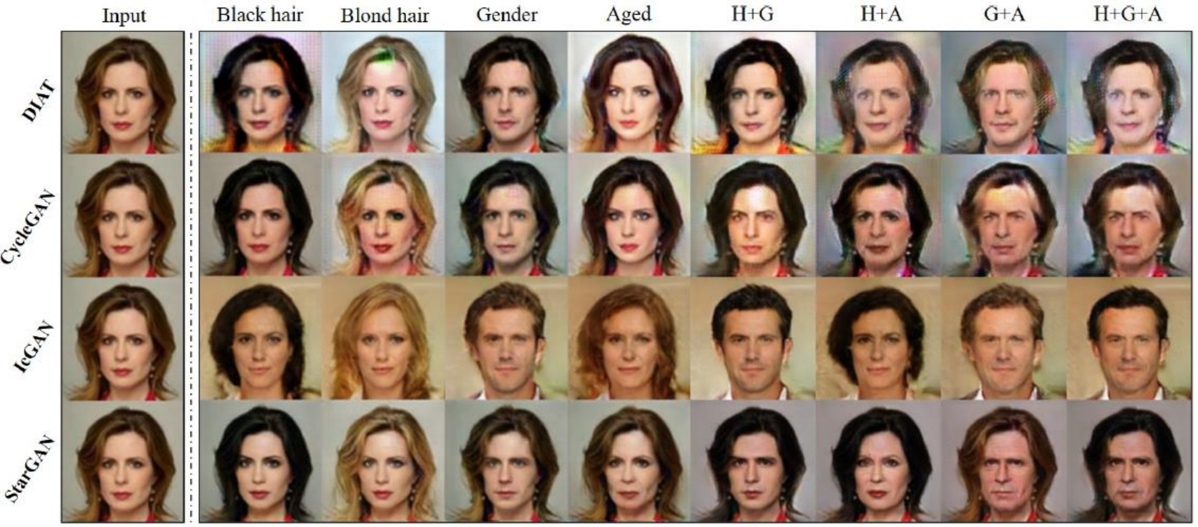
[MobileFaceNet: 1.5M]

# StarGAN

[Choi et al. CVPR 2018 (oral)]








| Method | Hair color | Gender | Aged |
|---|---|---|---|
| DIAT | 9.3% | 31.4% | 6.9% |
| CycleGAN | 20.0% | 16.6% | 13.3% |
| IcGAN | 4.5% | 12.9% | 9.2% |
| StarGAN | **66.2%** | **39.1%** | **70.6%** |

Table 1. AMT perceptual evaluation for ranking different models on a single attribute transfer task. Each column sums to 100%.

| Method | Classification error | # of parameters |
|---|---|---|
| DIAT | 4.10 | 52.6M × 7 |
| CycleGAN | 5.99 | 52.6M × 14 |
| IcGAN | 8.07 | 67.8M × 1 |
| StarGAN | **2.12** | **53.2M** × **1** |
| Real images | 0.45 | - |

| Method | H+G | H+A | G+A | H+G+A |
|---|---|---|---|---|
| DIAT | 20.4% | 15.6% | 18.7% | 15.6% |
| CycleGAN | 14.0% | 12.0% | 11.2% | 11.9% |
| IcGAN | 18.2% | 10.9% | 20.3% | 20.3% |
| StarGAN | **47.4%** | **61.5%** | **49.8%** | **52.2%** |

Table 2. AMT perceptual evaluation for ranking different models on a multi-attribute transfer task. H: Hair color; G: Gender; A: Aged.

# LaRva: Language Representation by Clova

Clova

- Giant-scale general-purpose language model by improving BERT
- Not Multi-lingual BERT but LaRva

**[WikiSQL]**

| Model | Dev logical form accuracy | Dev execution accuracy | Test logical form accuracy | Test execution accuracy | Uses execution |
|---|---|---|---|---|---|
| SQLova +Execution-Guided Decoding (Hwang 2019) | 84.2 | 90.2 | 83.6 | 89.6 | Inference |
| IncSQL +Execution-Guided Decoding (Shi 2018) | 51.3 | 87.2 | 51.1 | 87.1 | Inference |
| Execution-Guided Decoding (Wang 2018) | 76.0 | 84.0 | 75.4 | 83.8 | Inference |
| SQLova (Hwang 2019) | 81.6 | 87.2 | 80.7 | 86.2 | |
| IncSQL (Shi 2018) | 49.9 | 84.0 | 49.9 | 83.7 | |
| MQAN (unordered) (McCann 2018) | 76.1 | 82.0 | 75.4 | 81.4 | |
| MQAN (ordered) (McCann 2018) | 73.5 | 82.0 | 73.2 | 81.4 | |
| Coarse2Fine (Dong 2018) | 72.5 | 79.0 | 71.7 | 78.5 | |

**[KorQuAD]**

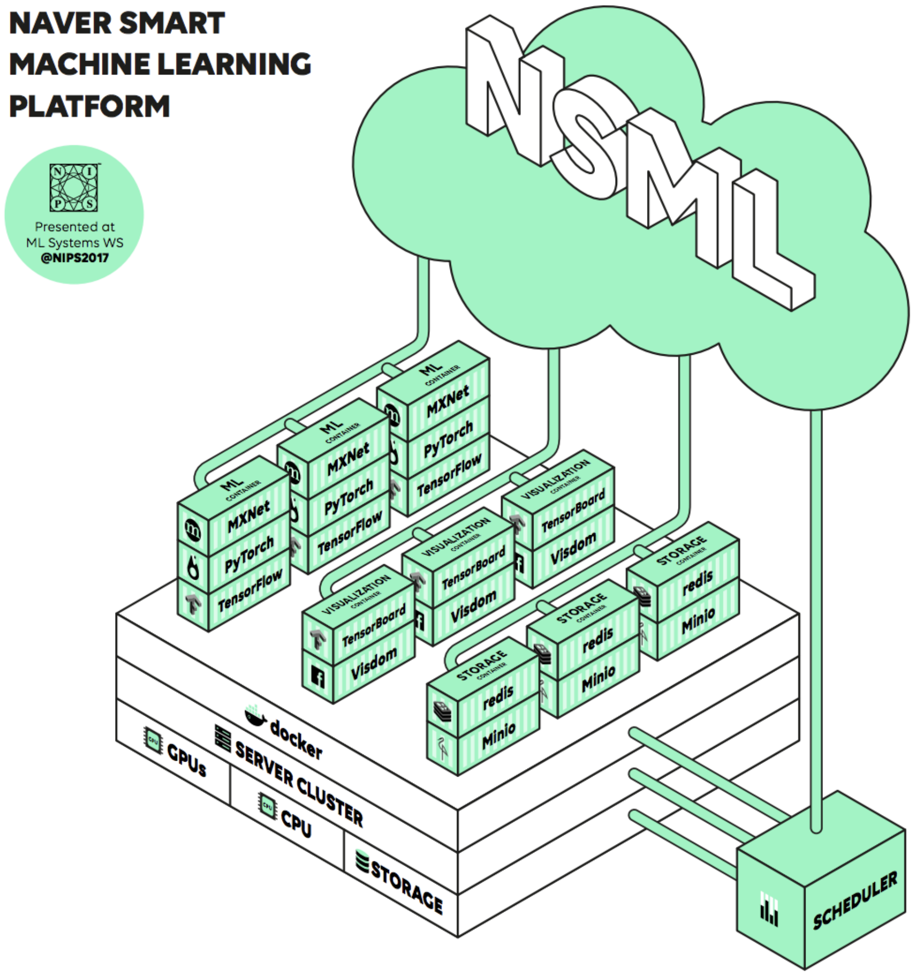
Leaderboard

| Rank | Reg. Date | Model | EM | F1 |
|---|---|---|---|---|
| - | 2018.10.17 | **Human Performance** | 80.17 | 91.20 |
| 1 | 2019.03.13 | BERT-Kor (single)<br>Clova AI LPT Team | **83.50** | **92.41** |
| 2 | 2019.01.30 | BERT LM fine-tuned (single) + KHAIII<br>Kakao NLP Team | 83.32 | 92.10 |
| 3 | 2019.01.24 | BERT LM fine-tuned (single) + KHAIII<br>Kakao NLP Team | 82.14 | 91.85 |

Clova

# NSML

NSML: ML research platform that enables you to focus on your model!!




[Sung et al. MLSYS 2017@NIPS 2017]

# Easy One-Liner CLI

- Dataset registration

```
/app/examples/09_NMT$ nsml dataset push NMT_EN_KR ./nmt_en_kr
```

- Train

```
/app/examples/09_NMT$ nsml run -d NMT_EN_KR
Session clair/NMT_EN_KR/1 is running
```

- Serve (Inference)

```
/app/examples$ echo Hello | nsml infer clair/NMT_EN_KR/1/12
안 녕 하 세 요
```

# NSML

Clova

- Support any kind of GPU clusters
- Automated GPU allocation / release
- Support most deep learning libraries
- Docker-based isolated research environment
- Easy-to-use CLI and Web interfaces
- Effective visualization
- Leaderboard
- Jupyter and AutoML

[Sung et al. MLSYS 2017@NIPS 2017]

# 네이버 스마트 머신러닝 플랫폼

## 쉽고 편리한 사용

**클라우드 기반**으로 즉시 사용이 가능하고,
데이터 관리가 편리합니다

## 효율적 자원 활용

**GPU Clustering** 및 **Scheduling** 를
통해 자원을 효과적으로 배분합니다

## 다양한 실험과 인사이트

**병렬학습** 및 **형상관리**, **Visualization** 을 통해
빠르게 인사이트를 얻습니다

## 효과적 협업과 투명한 관리

**문제를 공유**할 수 있고 전체 학습 및 자원 현황을
실시간 확인할 수 있습니다

## 모델 성능 최적화

**AutoML**을 통해 자동으로 파라미터를
튜닝하고 성능을 최적화합니다

## 빠르고 효율적인 데이터 레이블링

**Active Learning** 을 통해
성능 향상에 중요한 데이터를 선별하고,
자동으로 데이터를 레이블링 합니다

# AutoML in NSML

NSML

+

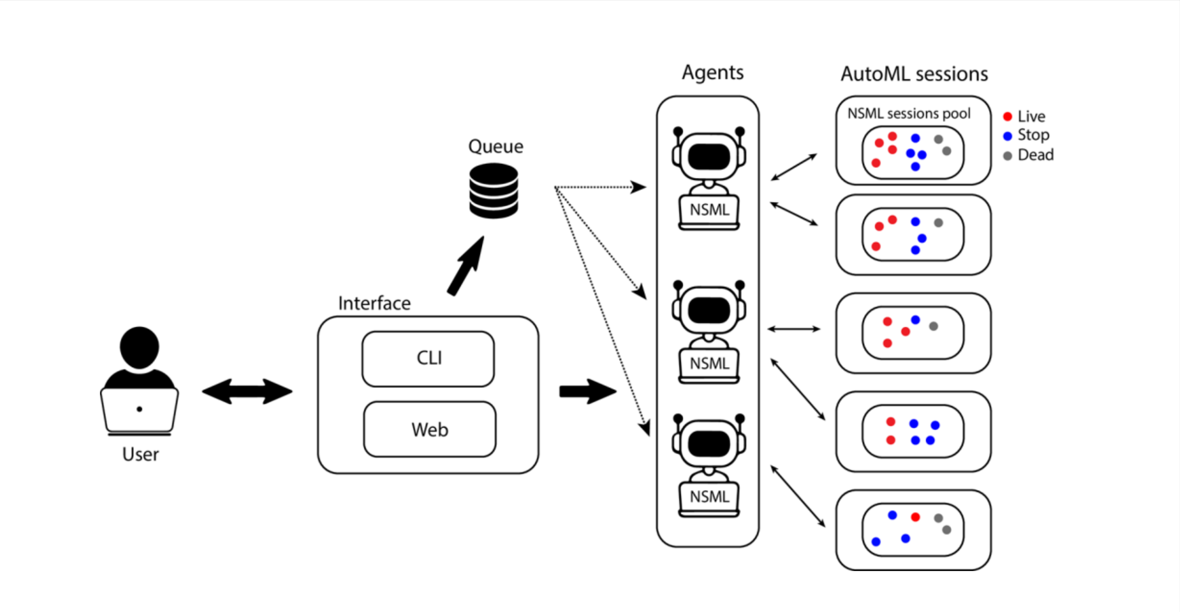
jupyter

AutoML

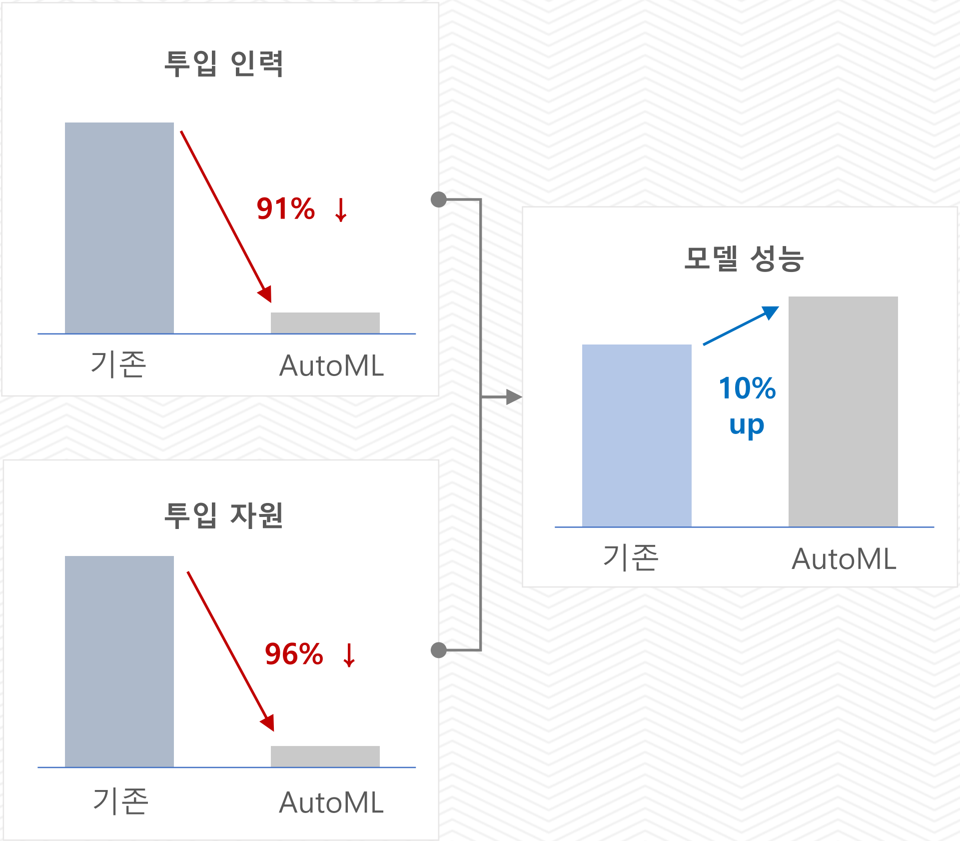
# (5) AutoML을 통해 더 빠르게, 더 좋은 학습 결과를 얻을 수 있습니다

### 파라미터 튜닝 자동화

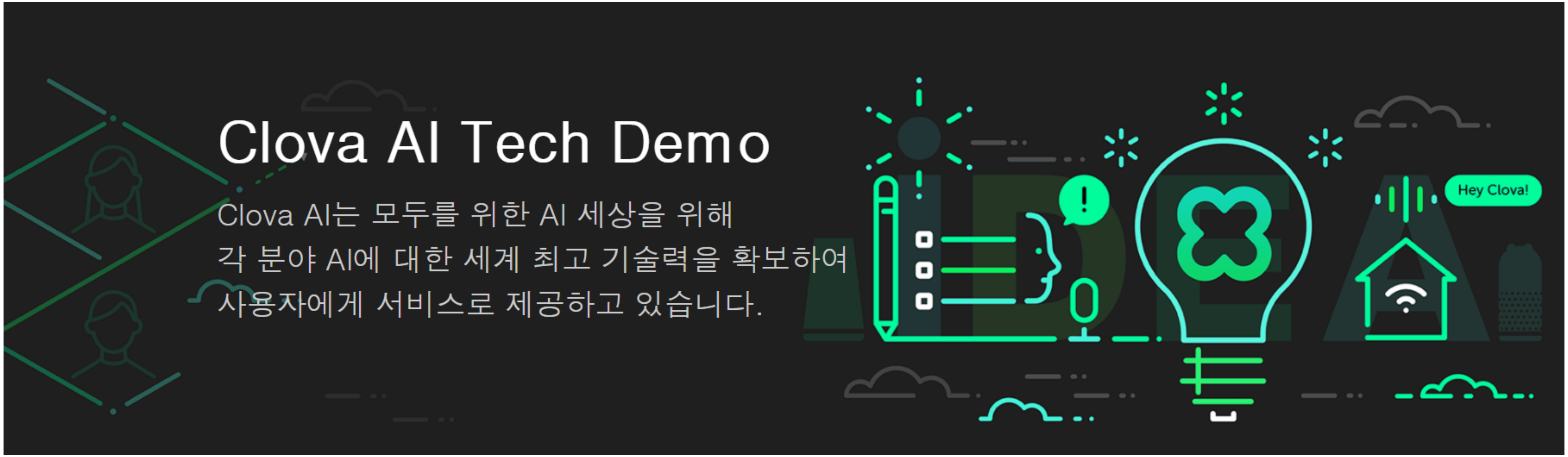
- 간단한 조건 세팅으로 Auto ML을 통해 자동으로 Parameter 를 조정하고 최적 모델 탐색
- 각 모델의 Parameter 조합 및 성능은 그래프를 통해 비교 확인
- 설정에 따른 성능 추이를 분석하고, 탐색 범위를 좁혀가며 추가적인 실험 진행




AutoML 활용 사례



# Clova AI Tech Demo




https://clova.ai/techdemo

# We can meet here

ICLR | 2019
Seventh International Conference on Learning Representations

ICML | 2019
Thirty-sixth International Conference on Machine Learning

CVPR
LONG BEACH CALIFORNIA
June 16-20, 2019

SIGIR 2019

ACL 2019 Florence
ANNUAL MEETING July 28th - August 2nd
of the Association for Computational Linguistics

ICCV 2019
Seoul, Korea

GRAZ – AUSTRIA
SEPTEMBER 15th – 19th 2019
INTERSPEECH 2019

EMNLP-IJCNLP 2019

The ACM Conference Series on
Recommender Systems

MobiSys2019
The 17th ACM international Conference on Mobile Systems, Applications and Services

Clova
Developers

# Research Opportunities

# Clova AI
# World-class Achievement

86%
(recall@1 vs Alibaba 79%)

Product Image Search

75% (1.6M)
(vs 72.1% with 1.7M, Intel Lab)

Lightweight (detection)

2%
(err rate vs others 5+

Four-hour Recoding Voice

99%

Pose Tracking
VLive ...


83%
(WIKISQL SOTA 75%)


80%
(Eng recg. vs google 54

Optical Character Recognitio

# Selected Publication List of Clova AI Since 2018

1. Sung et al. NSML: A Machine Learning Platform That Enables You to Focus on Your Models, MLSYSWS@NIPS 2017
2. Seo et al. Neural Speed Reading via Skim-RNN, ICLR 2018.
3. Choi et al. StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation, CVPR 2018.
4. Afouras et al. Deep Lip Reading: a comparison of models and an online application, Interspeech 2018.
5. Chung et al. VoxCeleb2: Deep Speaker Recognition, Interspeech 2018.
6. Afouras et al. The Conversation: Deep Audio Visual Speech Enhancement, Interspeech 2018.
7. Lee et al. Acoustic modeling using adversarially trained variational recurrent neural network for speech synthesis, Interspeech 2018.
8. Hwang et al. A Unified Framework for the Generation of Glottal Signals in Deep Learning-Based Parametric Speech Synthesis Systems, Interspeech 2018.
9. Park et al. Representation Learning of Music Using Artist Labels, ISMIR 2018.
10. Lee et al. Unsupervised holistic image generation from key local patches, ECCV 2018.
11. Kim et al. Multimodal Dual Attention Memory for Video Story Question Answering, ECCV 2018.
12. Seo et al. Phrase-Indexed Question Answering: A New Challenge for Scalable Document Comprehension, EMNLP 2018.
13. Lee et al. Answerer in Questioner's Mind: Information Theoretic Approach to Goal-Oriented Visual Dialog, NeurIPS 2018.
14. Song et al. Hierarchical Context enabled Recurrent Neural Network for Recommendation, AAAI 2019.
15. Park et al. Adversarial Dropout for Recurrent Neural Networks, AAAI 2019.
16. Park et al. Paraphrase Diversification using Counterfactual Debiasing, AAAI 2019.
17. Heo et al. Knowledge Distillation with Adversarial Samples Supporting Decision Boundary, AAAI 2019.
18. Heo et al. Knowledge Transfer via Distillation of Activation Boundaries Formed by Hidden Neurons, AAAI 2019.
19. Oh et al. Modeling Uncertainty with Hedged Instance Embeddings, ICLR 2019.
20. Gu et al. DialogWAE: Multimodal Response Generation with Conditional Wasserstein Auto-Encoder, ICLR 2019.
21. Lee et al. Large-Scale Answerer in Questioner's Mind for Visual Dialog Question Generation, ICLR 2019.
22. Kim et al. Curiosity-Bottleneck: Exploration By Distilling Task-Specific Novelty, ICML 2019.
23. Chung et al. Perfect match: Improved cross-modal Embeddings for audio-visual synchronisation, ICASSP 2019
24. Baek et al. Character Region Awareness for Text Detection, CVPR 2019.
25. Seo et al. Real-Time Open-Domain Question Answering with Dense-Sparse Phrase Index, ACL 2019.
26. Chung et al. Who said that?: Audio-visual speaker diarisation of real-world meeting, Interspeech 2019.
27. Afouras et al. My lips are concealed: Audio-visual speech enhancement through obstructions, Interspeech 2019.
28. Yamamoto et al. Probability Density Distillation with Generative Adversarial Networks for High-quality Parallel Waveform Generation, Interspeech 2019.
29. Hwang et al. Parameter enhancement for MELP speech codec in noisy communication environment, Interspeech 2019.

32%

# WE ARE HIRING! Join Us!

## Positions

Research Scientist | AI Software Engineer | Research Internship | Global Residency

## Fields and Domains

All fields of AI from fundamental theories to practical applications

## Why should I Join Naver Clova?

Great Colleagues & Work Environment, Hands-on R&D Experience
Advisory: Kyunghyun Cho, Jun-Yan Zhu, Hannaneh Hajishirzi, and Jaegul Choo

## How to Apply?

clova-jobs@navercorp.com

# We are hiring 2019 Fall / 2020 Spring Global Residency

## Clova AI Research Recruiting

## 2020 SPRING GLOBAL RESIDENCY

# Send Your CV Today!

clova-jobs@navercorp.com

[Clova AI Github]

[Facebook Clova AI Research]

[Clova AI Research Web]