

PORTFOLIO

Lee Tae Bok

Multimedia Signal Processing Lab.
Department of Artificial Intelligence
Ajou University

About me



이태복 Tae Bok Lee

Computer Vision & Artificial Intelligence Researcher

Contact

- dolphin0104@gmail.com
- 010-4678-0053

Education

- 2023. 08.** **Ph.D. course in Artificial Intelligence**
Ajou University, Suwon, Korea
- 지도교수: 허용석
 - Lab: Multimedia Signal Processing Lab
 - Dissertation: *Accurate Blur Decomposition From a Single Image Using Conditional GANs*
- 2018. 08.** **B.S. in Electrical and Computer Engineering**
Ajou University, Suwon, Korea

Research Background

Computer Vision and Artificial Intelligence

- Image Restoration (deblurring, super resolution)
- Conditional image synthesis based on GANs
- Face detection and recognition

Achievements

▪ Publications

- **Tae Bok Lee** and Yong Seok Heo, "ABDGAN: Arbitrary Time Blur Decomposition Using Critic-Guided TripleGAN", Submitted to ICCV 2023
- **Tae Bok Lee**, Suhy Han, and Yong Seok Heo, "Continuous Facial Motion Deblurring," IEEE Access, vol. 10, no. 1, pp. 76079 - 76094, Jul. 2022. [[Paper](#)] [[Github](#)]
- **Tae Bok Lee**, Soo Hyun Jung, and Yong Seok Heo, "Progressive Semantic Face Deblurring," IEEE Access, vol. 8, no. 1, pp. 223548 - 223561, Oct. 2020. [[Paper](#)] [[Github](#)]
- **Tae Bok Lee** and Yong Seok Heo, "Single Image Super Resolution Using Convolutional Neural Networks for Noisy Images," The 11th International Conference on ICT Convergence (ICTC), 2020. [[Paper](#)]
- Suhy Han, **Tae Bok Lee**, and Yong Seok Heo, "Semantic-Aware Face Deblurring With Pixel-Wise Projection Discriminator," IEEE Access, vol. 11, no. 1, pp. 11587 - 11600, Feb. 2023. [[Paper](#)] [[Github](#)]
- Soo Hyun Jung, **Tae Bok Lee**, and Yong Seok Heo, "Deep Feature Prior Guided Face Deblurring," IEEE Winter Conference on Applications of Computer Vision (WACV), 4-8 Jan. 2022. [[Paper](#)]
- Chang Jong Shin, **Tae Bok Lee**, and Yong Seok Heo, "Dual Image Deblurring Using Deep Image Prior," Electronics, 10(17), 2045, Aug. 2021. [[Paper](#)]
- Suhy Han, **Tae Bok Lee**, and Yong Seok Heo, "Deep Image Prior for Super Resolution of Noisy Image," Electronics, 10(16), 2045, Aug. 2021. [[Paper](#)] [[Github](#)]

▪ Patents

- 허용석, 이태복, 한수지, 화소 단위 사영 판별자를 이용한 영상 노이즈 제거 방법 및 장치 (출원) (10-2023-0012395) (Jan, 2023)
- Yong Seok Heo, Soo Hyun Jung, Tae Bok Lee, METHOD AND APPARATUS FOR PROCESSING BLURRED IMAGE (등록) (11,488,288) (Nov, 2022)
- 허용석, 이태복, 한수지, 블러 이미지에 포함된 연속적인 이미지의 취득 방법 및 장치 (출원) (10-2022-0124693) (Sep, 2022)
- 허용석, 이태복, 정수현, 얼굴 영상의 점진적 디블러링 방법 및 장치 (등록) (10-2358355) (Jan, 2022)
- 허용석, 이태복, 저해상도 이미지의 영상 복원 방법 및 장치 (출원) (10-2020-0146356) (Nov, 2020)

▪ Participated Projects

- "딥러닝 기반 정밀 Semantic Map 추출 엔진 개발", 삼성 전자, 2018-2019.
- "UHD 방송 환경에서 콘텐츠에 대한 시청자의 반응 및 의도 기반 미디어 인터랙션 기술", 과학기술정보통신부, 2018-2020.
- "효율적인 얼굴 인식을 위한 딥러닝 기법 연구", 슈프리마, 2019-2020.
- "위성정보 융합 서비스 ICT 인력 양성 연구센터", ITRC, 2018-2023

Publications

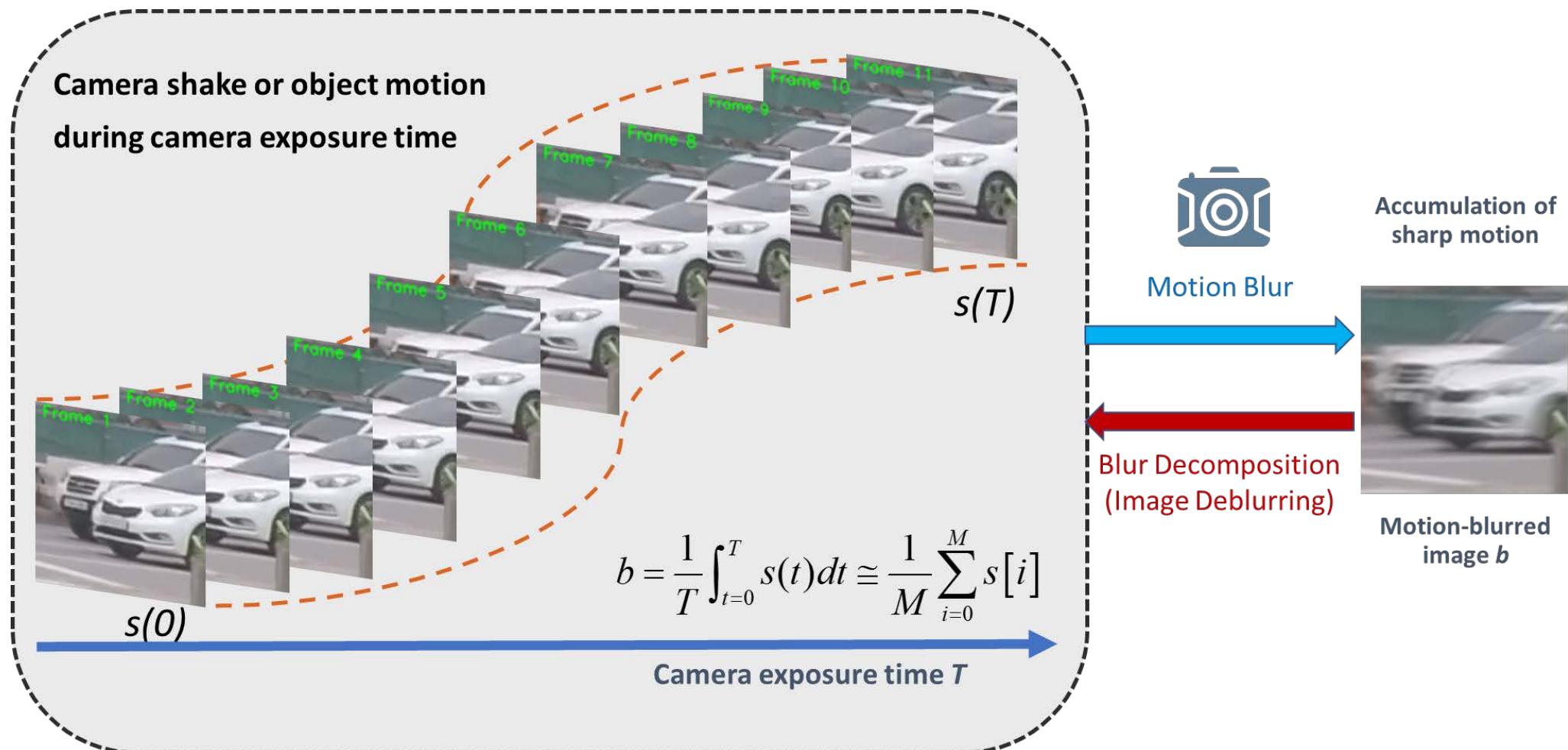
ABDGAN: Arbitrary Time Blur Decomposition Using Critic-Guided TripleGAN

Tae Bok Lee and Yong Seok Heo

Submitted to ICCV 2023

What Happened in the Motion Blurred Image?

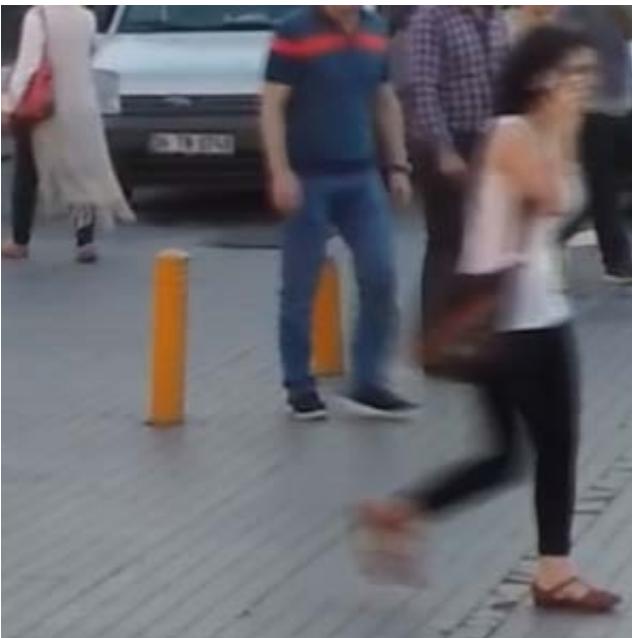
- Motion blur is an accumulation of continuous sharp motion.



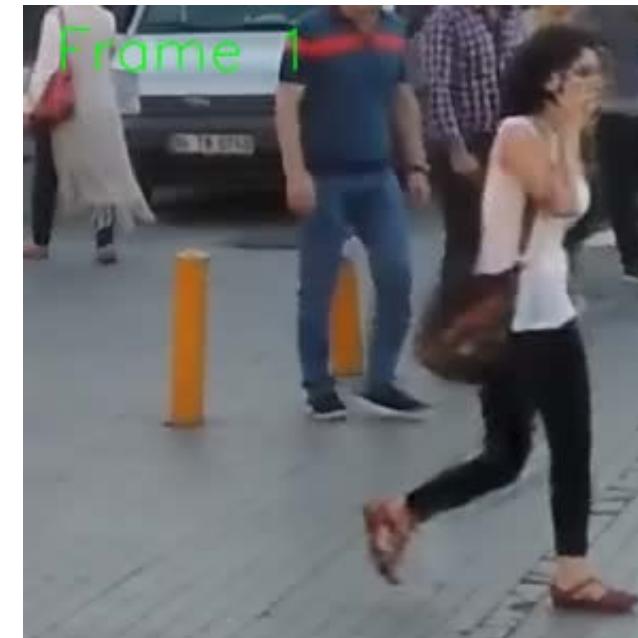
ABDGAN: Arbitrary Time Blur Decomposition Using Critic-Guided TripleGAN

Tae Bok Lee and Yong Seok Heo
Submitted to ICCV 2023

Our Goal : Finding Sharp Motion Latent in a Blurred Image



Blurred Image



Sharp Motion

ABDGAN: Arbitrary Time Blur Decomposition Using Critic-Guided TripleGAN

Tae Bok Lee and Yong Seok Heo
Submitted to ICCV 2023

Application

- Uncovering underlying dynamics of scene
 - sports, traffic monitoring, entertainment.
- Benefiting many real-world applications
 - segmentation, detection, and recognition



Blur + Detection*



DeBlur ** + Detection*



Ground truth + Detection*

* Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

** Kupyn, Orest, et al. "Deblurgan: Blind motion deblurring using conditional adversarial networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

ABDGAN: Arbitrary Time Blur Decomposition Using Critic-Guided TripleGAN

Tae Bok Lee and Yong Seok Heo

Submitted to ICCV 2023

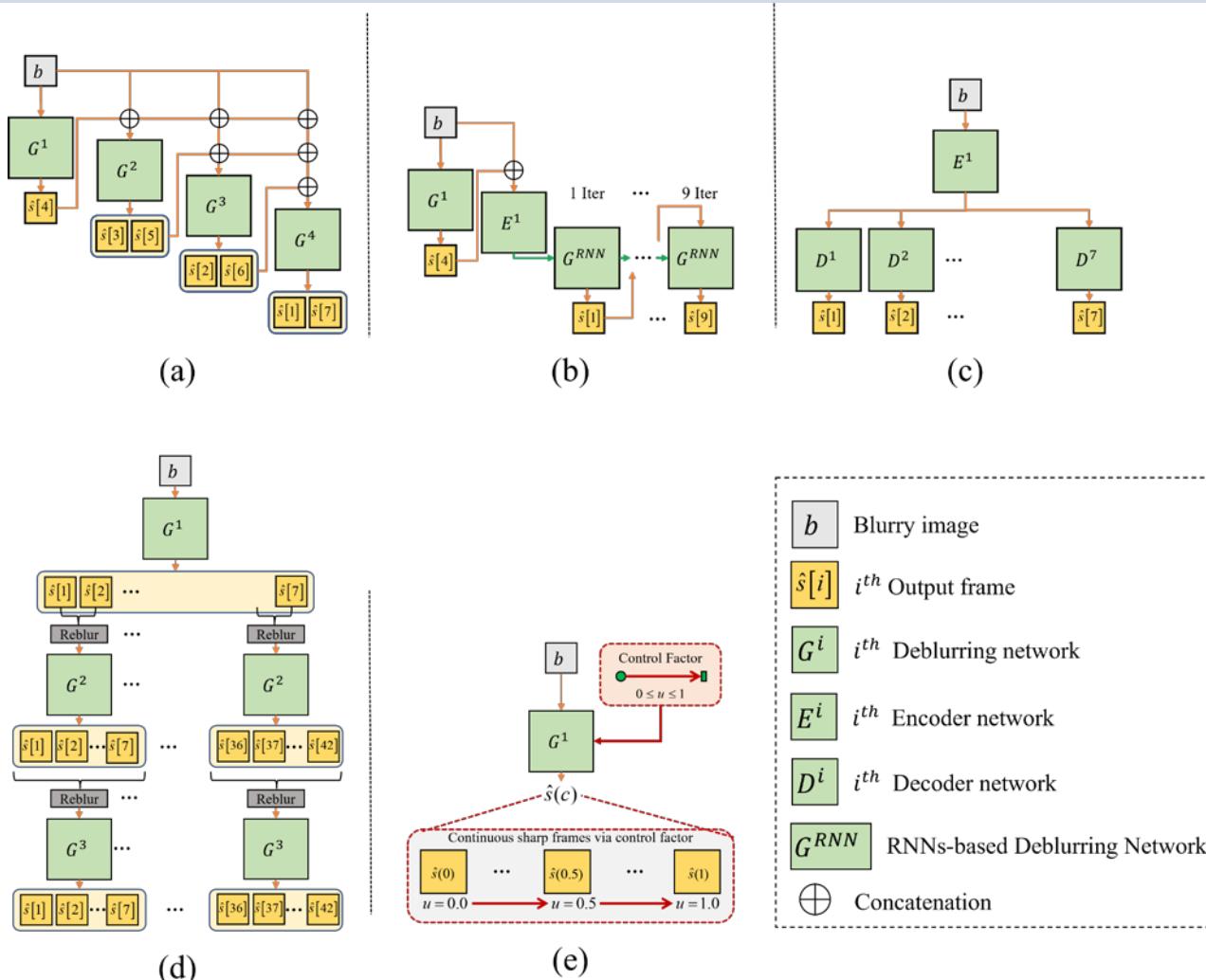
Single Image Deblurring

▪ Previous Works

- Single-to-Single deblurring
 - Focus on restoring a single sharp image
- Single-to-Video deblurring
 - Model is designed or trained to restore only a fixed number of frames.
 - Supervised learning.
 - Pairwise-order invariant loss.

▪ Our ABDGAN

- **Arbitrary Time Blur Decomposition:** Restore a *continuous sharp motion* from a single blurred image.
 - Our time-conditional deblurring network can output continuous sharp motion.
 - Semi-supervised learning to understand sharp motions not present in the training set.
 - Pairwise-order consistency loss for more accurate supervision

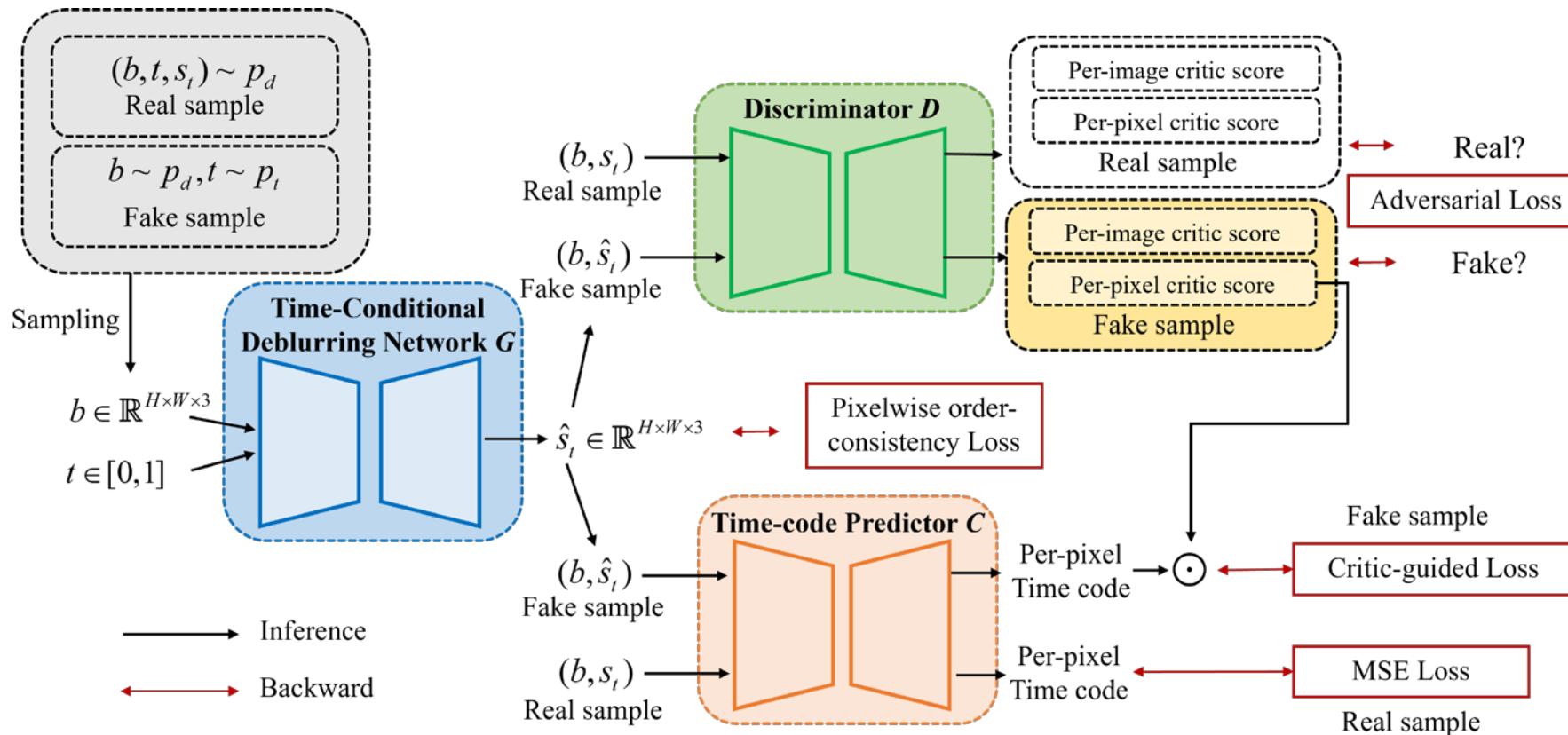


Comparison of single-to-video deblurring network architectures. The proposed method can restore continuous sharp motion with a single network. (a) Jin et al. (b) Purohit et al. (c) Argawal et al. (d) Zhang et al., and (e) proposed ABDGAN

ABDGAN: Arbitrary Time Blur Decomposition Using Critic-Guided TripleGAN

Tae Bok Lee and Yong Seok Heo
Submitted to ICCV 2023

Overall Framework of ABDGAN



- ABDGAN is proposed to learn the restoration of arbitrary time blur decomposition.
 - It learns to estimate sharp motion based on the input blurred image and an arbitrary time code.
 - It leverages a TripleGAN-based framework for semi-supervised learning.

ABDGAN: Arbitrary Time Blur Decomposition Using Critic-Guided TripleGAN

Tae Bok Lee and Yong Seok Heo

Submitted to ICCV 2023

Experiments

Comparison on GoPro dataset

	Methods	F_i	F_m	F_f	Avg.
PSNR (\uparrow)	Jin <i>et al.</i> [15]	24.04	26.98	23.12	24.71
	Argaw <i>et al.</i> [1]	<u>27.36</u>	29.84	<u>27.41</u>	27.90
	Zhang <i>et al.</i> [42]	18.78	31.05	18.75	22.87
	ABDGAN-GP	28.16	<u>30.16</u>	28.18	28.83
SSIM (\uparrow)	Jin <i>et al.</i> [15]	.833	.881	.810	.841
	Argaw <i>et al.</i> [1]	.794	.885	.793	.824
	Zhang <i>et al.</i> [42]	.688	.949	.672	.770
	ABDGAN-GP	.911	<u>.934</u>	.910	.918
LPIPS (\downarrow)	Jin <i>et al.</i> [15]	.308	.255	.323	.295
	Argaw <i>et al.</i> [1]	N/A	N/A	N/A	N/A
	Zhang <i>et al.</i> [42]	.428	<u>.192</u>	.442	.354
	ABDGAN-GP	.205	.149	.207	.187

Comparison on B-Aist++ dataset

	Methods	F_i	F_m	F_f	Avg.
PSNR (\uparrow)	Zhong <i>et al.</i> [43]	19.42	30.66	19.46	23.18
	ABDGAN-BA	21.50	31.38	21.53	24.80
SSIM (\uparrow)	Zhong <i>et al.</i> [43]	.846	.954	.846	.882
	ABDGAN-BA	.877	.957	.874	.903
LPIPS (\downarrow)	Zhong <i>et al.</i> [43]	.157	.076	.155	.129
	ABDGAN-BA	.126	.068	.126	.107

[1] M. Jin G. Meishvili and P. Favaro "Learning to extract a video sequence from a single motion-blurred image" Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) pp. 6334-6342 Jun. 2018.

[15] D. M. Argaw J. Kim F. Rameau C. Zhang and I. S. Kweon "Restoration of video frames from a single blurred image with motion understanding" Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW) pp. 701-710 Jun. 2021.

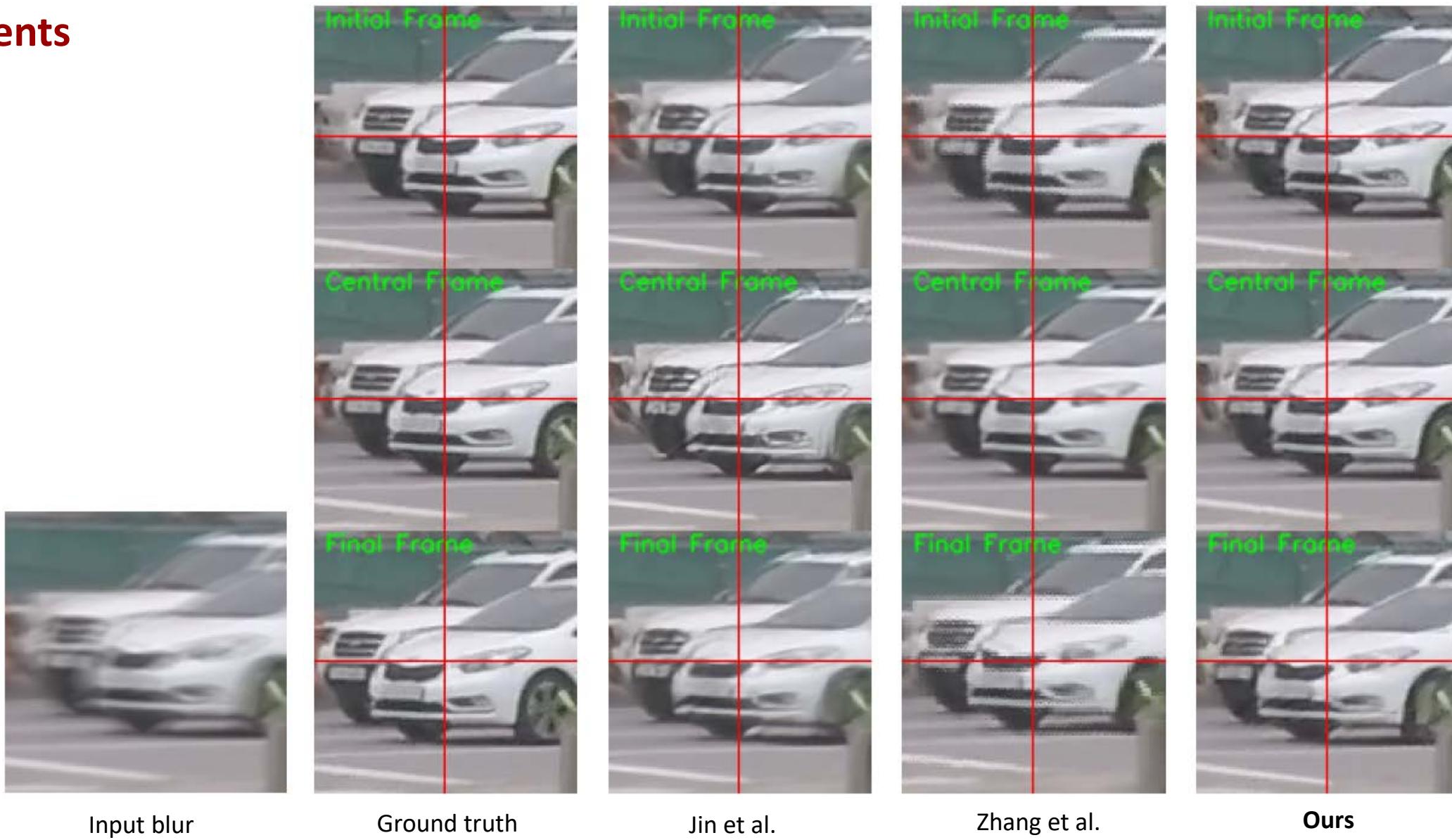
[42] Zhang, Youjian, et al. "Exposure trajectory recovery from motion blur." IEEE Transactions on Pattern Analysis and Machine Intelligence 44.11 (2021): 7490-7504.

[43] Zhong, Zhihang, et al. "Animation from blur: Multi-modal blur decomposition with motion guidance." Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XIX, Cham: Springer Nature Switzerland, 2022.

ABDGAN: Arbitrary Time Blur Decomposition Using Critic-Guided TripleGAN

Tae Bok Lee and Yong Seok Heo
Submitted to ICCV 2023

Experiments



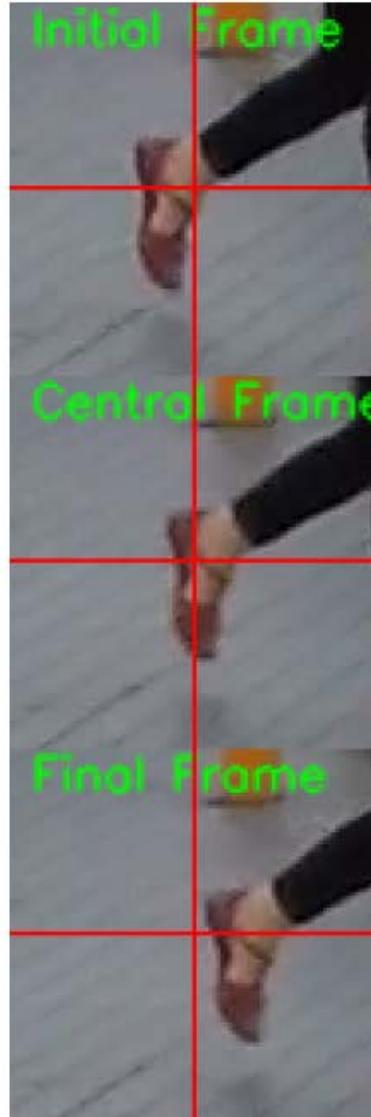
ABDGAN: Arbitrary Time Blur Decomposition Using Critic-Guided TripleGAN

Tae Bok Lee and Yong Seok Heo
Submitted to ICCV 2023

Experiments



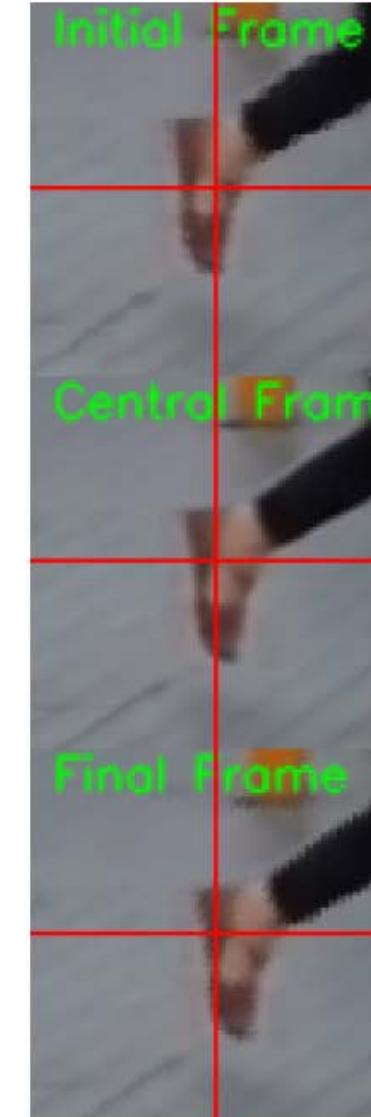
Input blur



Ground truth



Jin et al.



Zhang et al.



Ours

ABDGAN: Arbitrary Time Blur Decomposition Using Critic-Guided TripleGAN

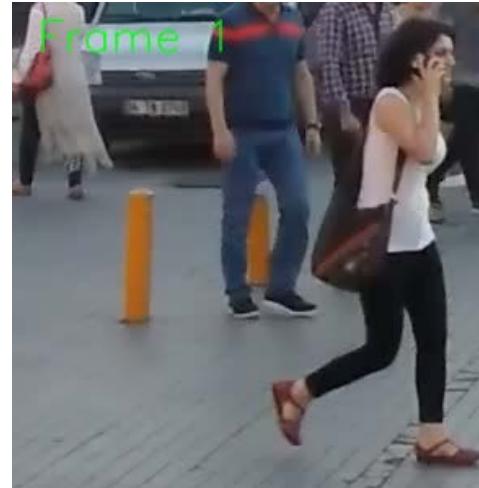
Tae Bok Lee and Yong Seok Heo

Submitted to ICCV 2023

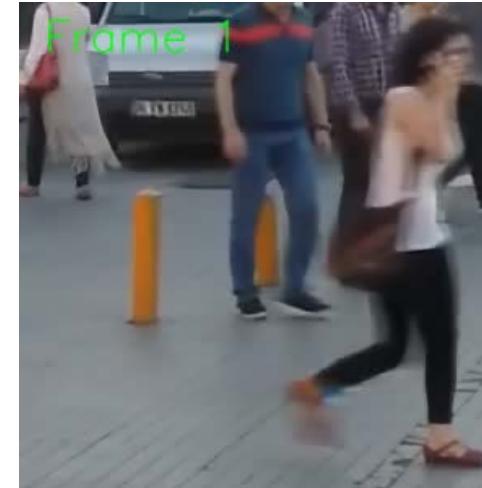
Experiments



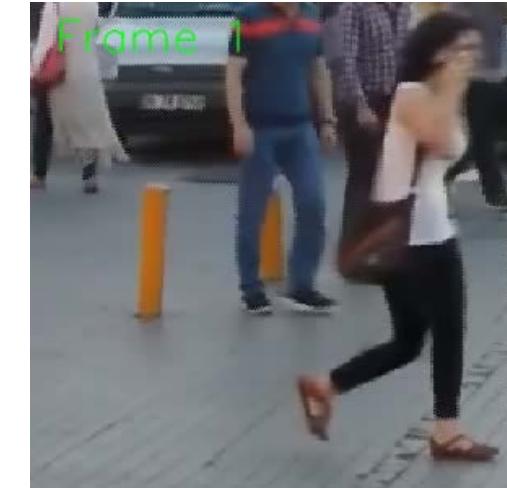
Input



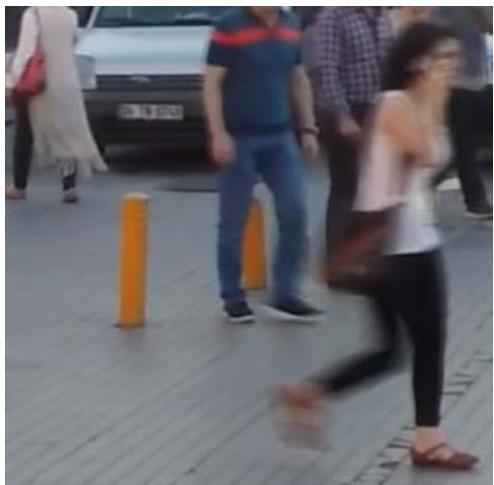
GT 11 frames



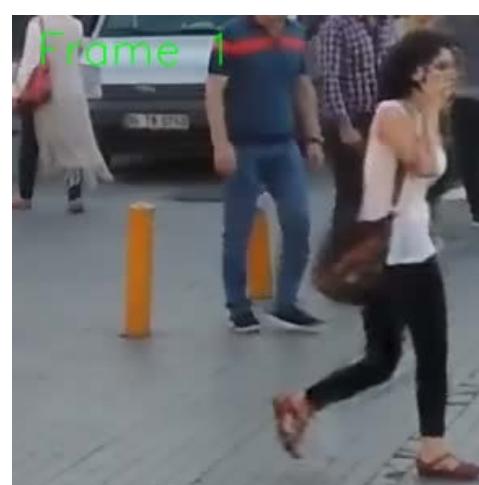
Jin 7 frames



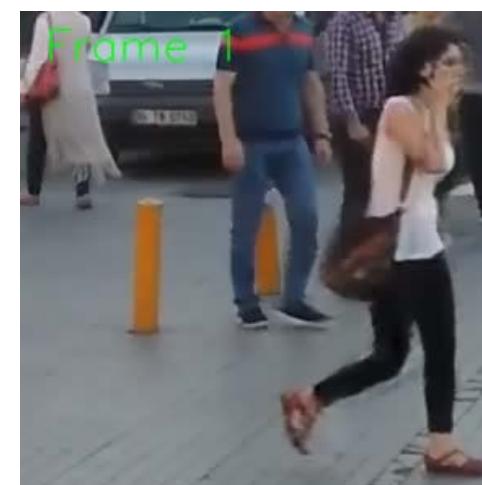
Zhang 15 frames



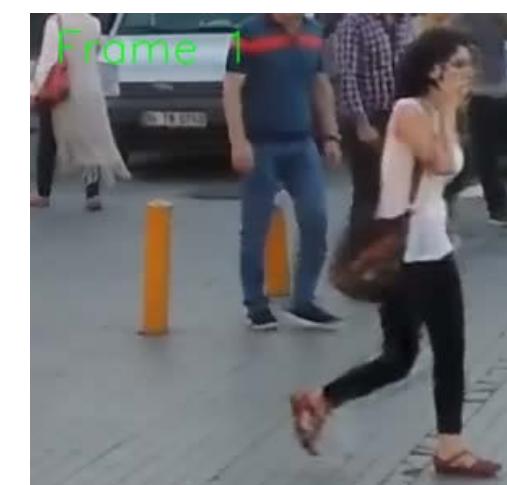
Magnified part of Input



Ours 11 frames



Ours 7 frames



Ours 15 frames

ABDGAN: Arbitrary Time Blur Decomposition Using Critic-Guided TripleGAN

Tae Bok Lee and Yong Seok Heo
Submitted to ICCV 2023

Experiments



Input



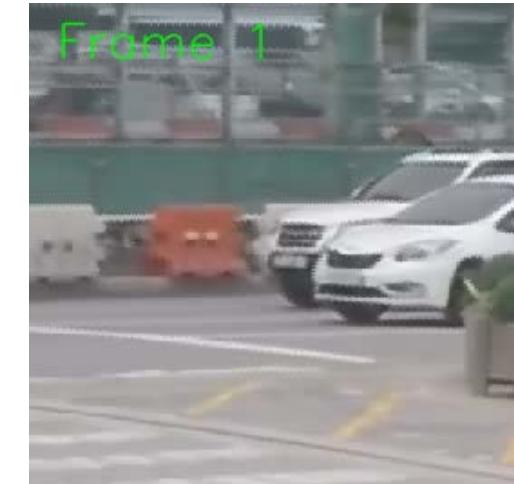
Magnified part of Input



GT 11 frames



Jin 7 frames



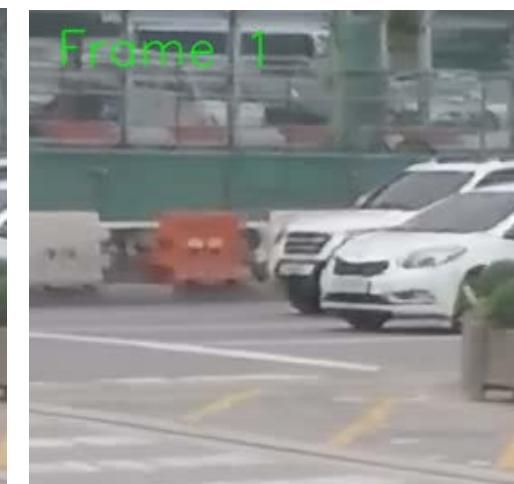
Zhang 15 frames



Ours 11 frames



Ours 7 frames



Ours 15 frames

ABDGAN: Arbitrary Time Blur Decomposition Using Critic-Guided TripleGAN

Tae Bok Lee and Yong Seok Heo
Submitted to ICCV 2023

Experiments



Input



GT 7 frames



Zhong 7 frames



Ours 7 frames

Continuous Facial Motion Deblurring

Tae Bok Lee, Suju Han, and Yong Seok Heo

IEEE Access, 2022

Our Goal: Continuous Facial Motion Deblurring



Blurry Face

Restoration of Continuous Facial Motion

▪ Previous Works

- Face deblurring methods restore only a single face image.
 - Existing single-to-video deblurring cannot generalize well on face images.

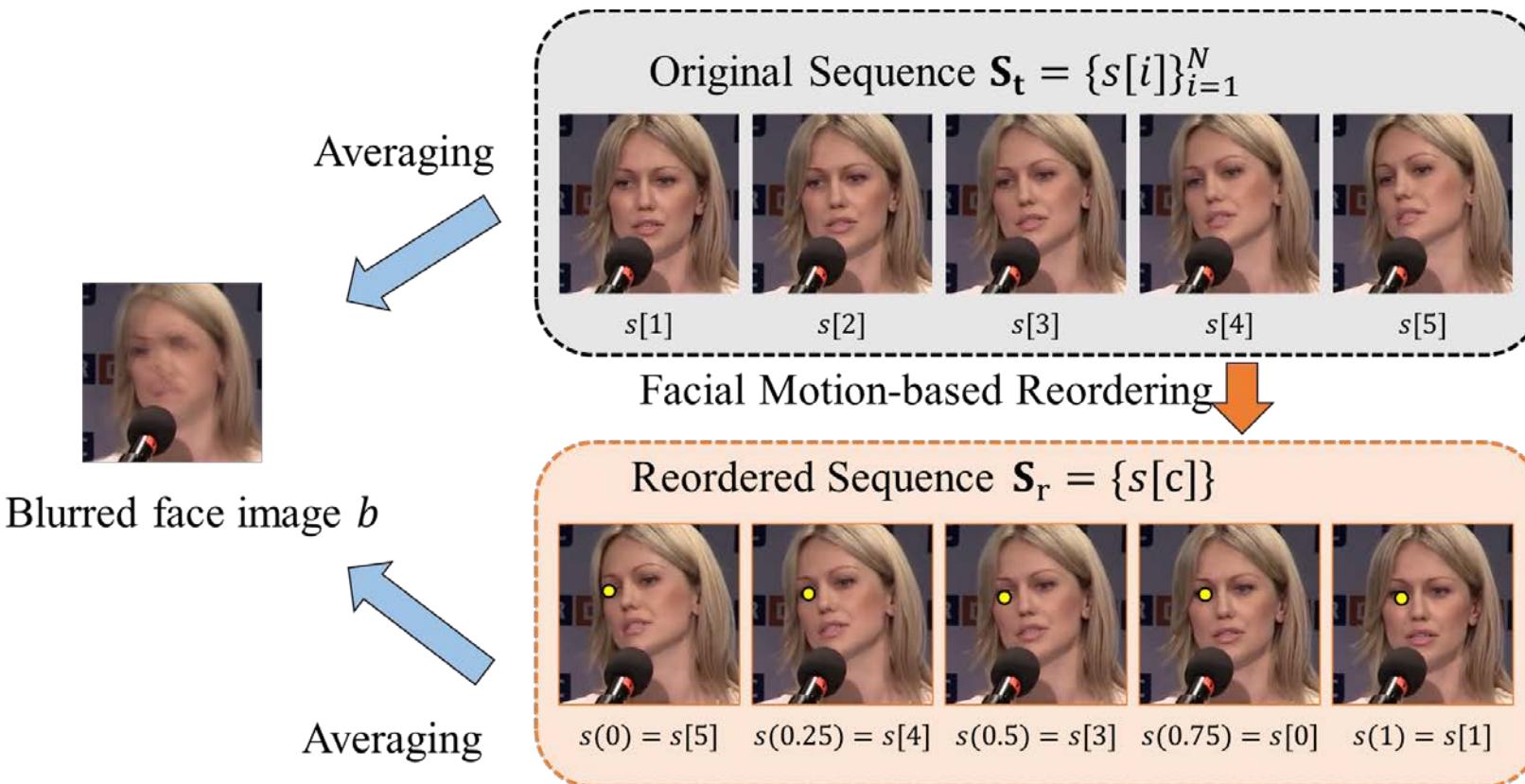
▪ Our CFMD-GAN

- Restore a continuous facial motion using a **control factor**
 - **Facial Motion-based Reordering (FMR)** : alleviate the difficulty of temporal ambiguity.
 - **Control-Adaptive (ContAda) block** : focus on important feature
 - [\[Paper\]](#) [\[Github\]](#)

Continuous Facial Motion Deblurring

Tae Bok Lee, Suhy Han, and Yong Seok Heo
IEEE Access, 2022

Proposed Facial Motion-based Reordering (FMR)



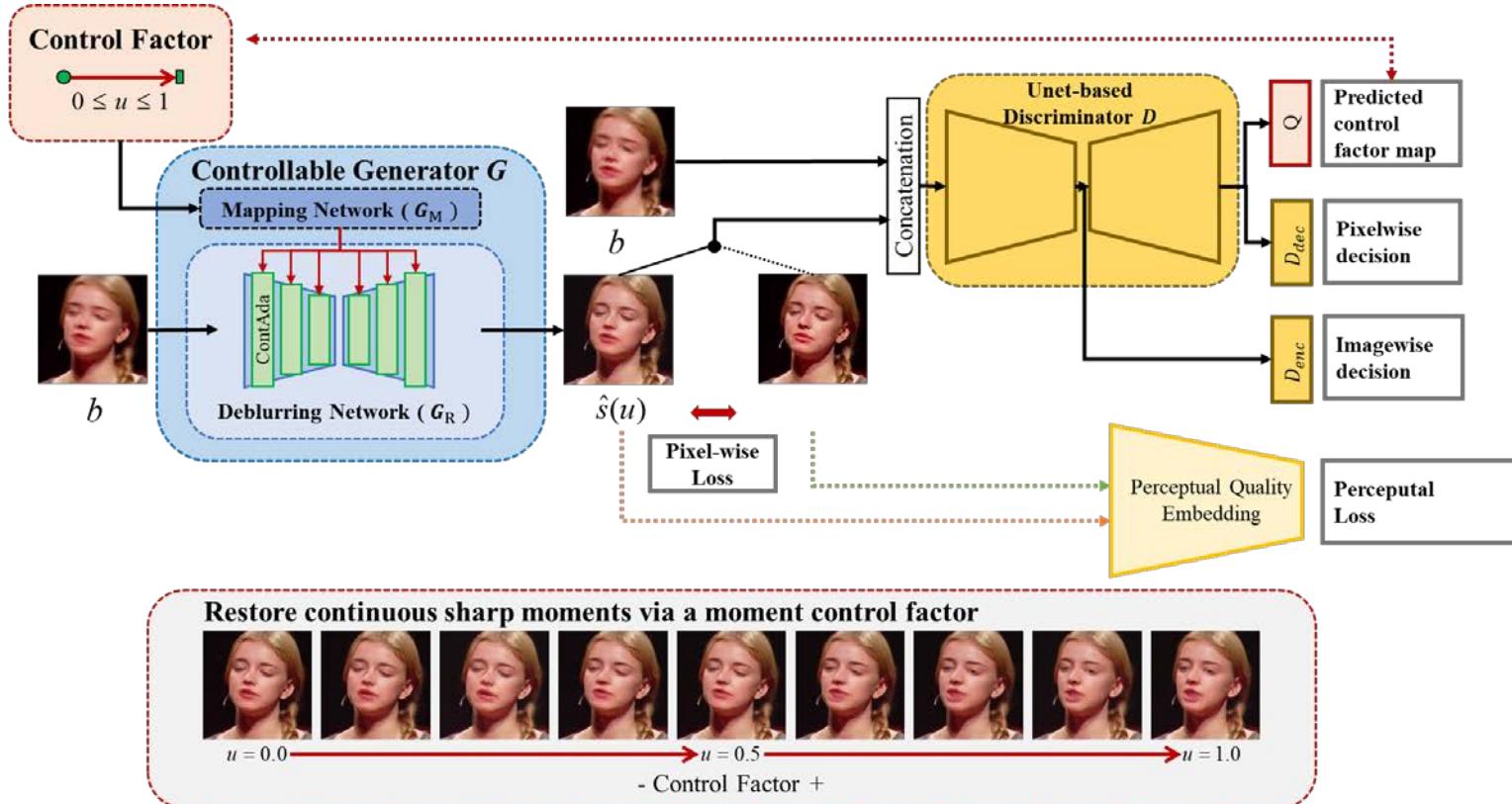
- Rearrange the original sequence based on the position of the left eye.
 - Network learns more generalized order => converge well
 - Network naturally learns the facial characteristic (left eye)

Continuous Facial Motion Deblurring

Tae Bok Lee, Suhy Han, and Yong Seok Heo

IEEE Access, 2022

Overall Framework of Our CFMD-GAN



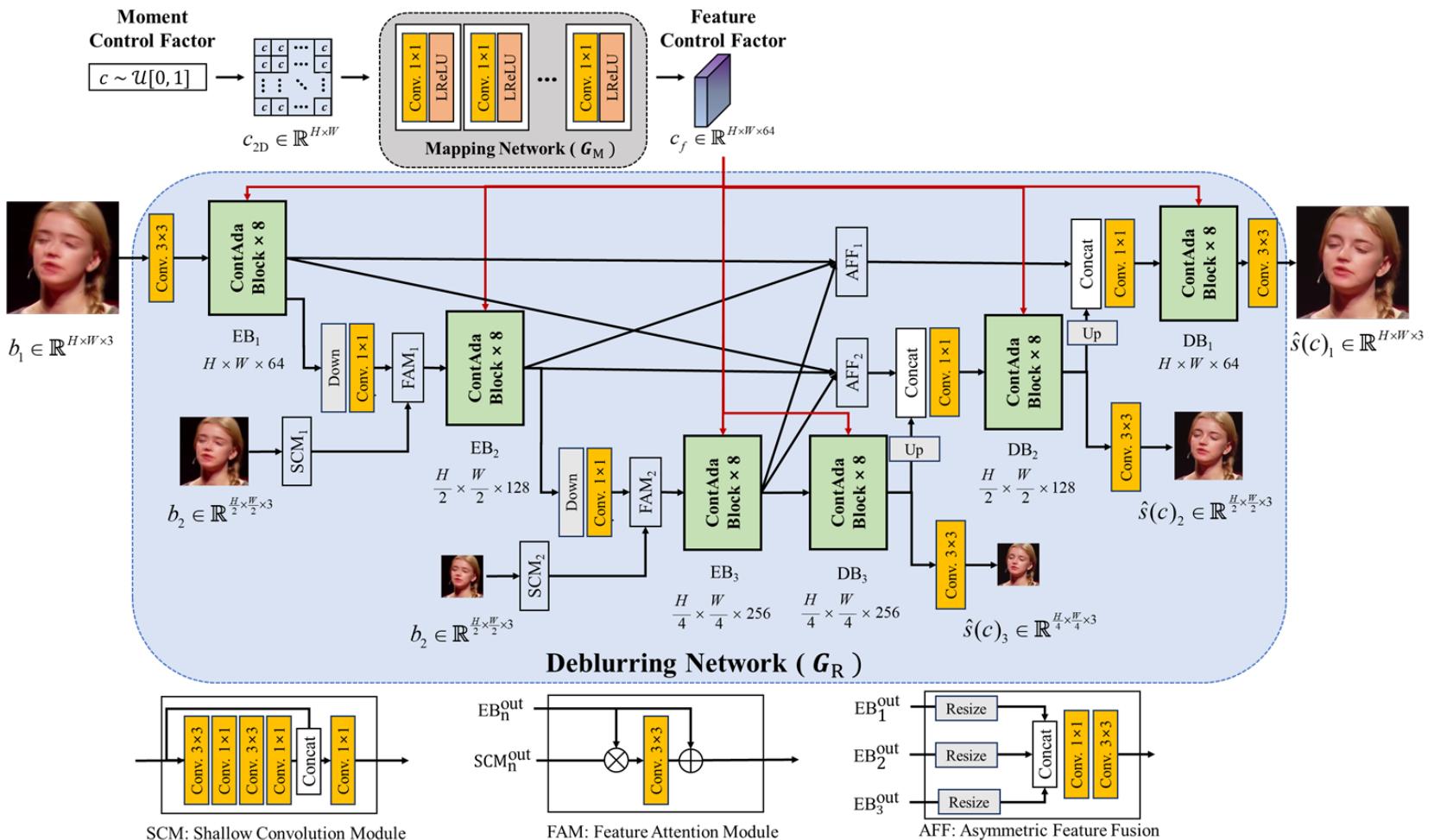
- Controllable blur decomposition framework based on **Conditional GAN**
- Restore continuous frames with a single network by varying a moment **control factor**.
- Does not require architectural changes and retraining process

Continuous Facial Motion Deblurring

Tae Bok Lee, Suhy Han, and Yong Seok Heo

IEEE Access, 2022

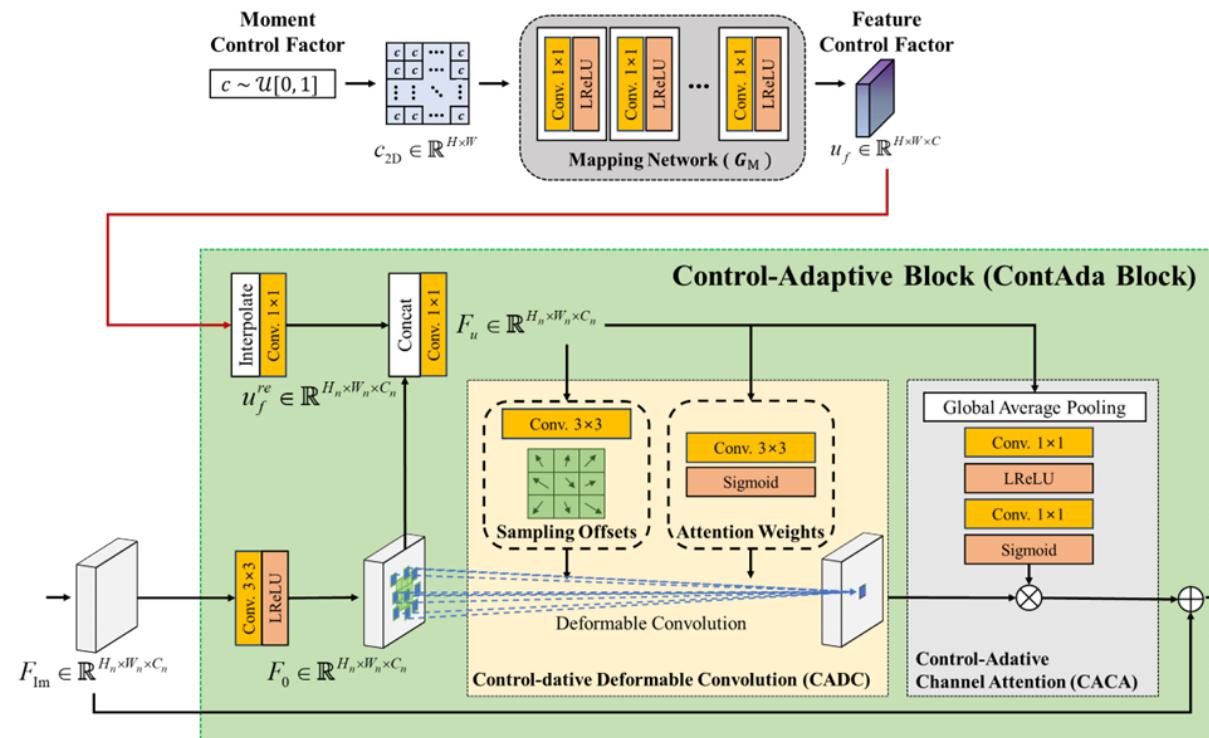
Generator (Deblurring Network) Architecture



Continuous Facial Motion Deblurring

Tae Bok Lee, Suhy Han, and Yong Seok Heo
IEEE Access, 2022

Control-Adaptive (ContAda) Block



▪ Control Adaptive Deformable Convolution (CADC)

- Learns **deformable offsets** and **attention weights**
- Allows network to better handle object deformations and variations in scale and orientation

▪ Control Adaptive Channel Attention (CACA)

- Learns **channel-wise weights** using image feature map and **control factor**
- Allows network to learn important channel-wise feature according to the control factor.

Continuous Facial Motion Deblurring

Tae Bok Lee, Suhy Han, and Yong Seok Heo
IEEE Access, 2022

Experiments

Quantitative Comparison with Single-to-Single General Deblurring Methods

Methods	PSNR (\uparrow)	SSIM (\uparrow)	LPIPS (\downarrow)	FID (\downarrow)	ArcFace (\downarrow)
Nah <i>et al.</i> [16]	31.4144	0.9232	0.0935	13.9722	1.1250
SRN [46]	32.1485	0.9249	0.0930	11.6292	1.1488
DMPHN [48]	33.1797	0.9284	0.0847	13.0407	1.0338
DMPHN*	33.8182	0.9345	0.0916	14.3071	1.0126
MIMO [49]	34.0372	0.9350	0.0795	7.8606	1.0205
MIMO*	34.8496	0.9401	<u>0.0794</u>	<u>7.2459</u>	<u>0.9918</u>
CFMD-GAN	<u>34.2684</u>	<u>0.9362</u>	0.0697	5.1448	0.9338

Quantitative Comparison with Single-to-Single Face Deblurring Methods

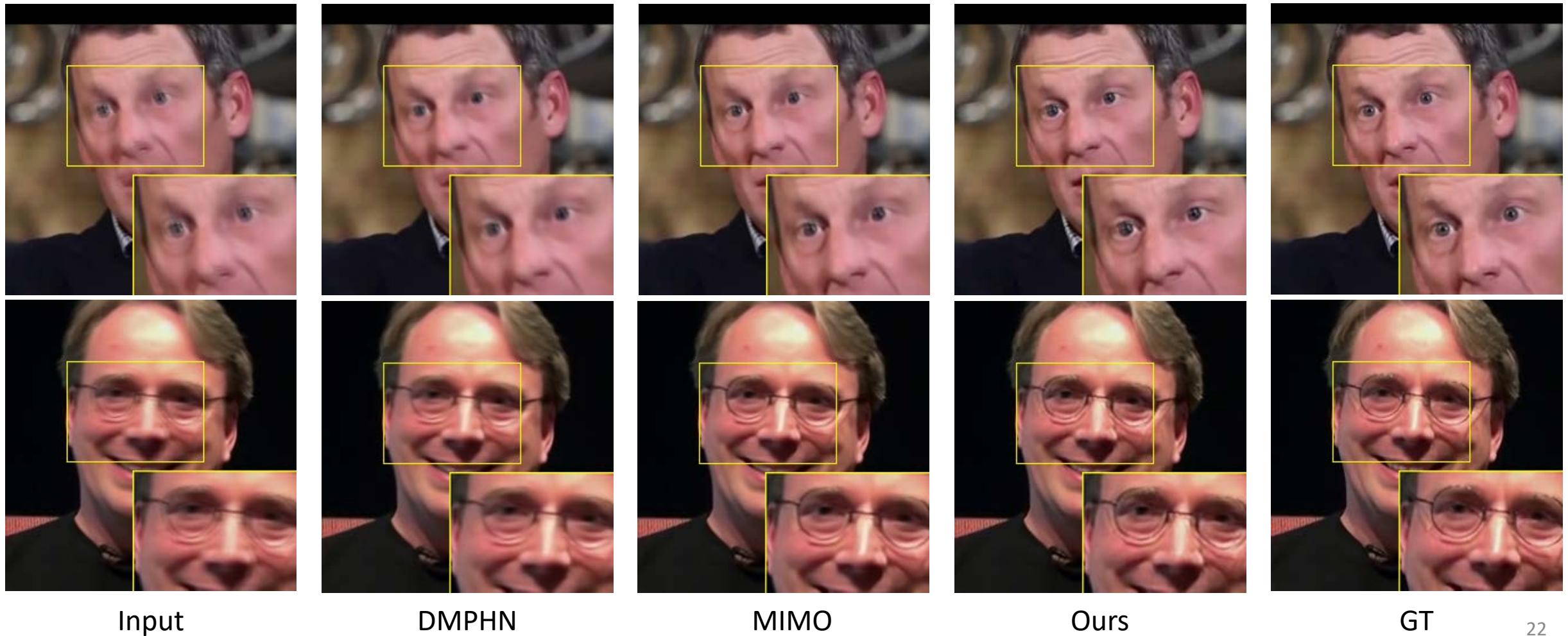
Methods	PSNR (\uparrow)	SSIM (\uparrow)	LPIPS (\downarrow)	FID (\downarrow)	ArcFace (\downarrow)
Shen <i>et al.</i> [18]	23.1795	0.6873	0.2310	78.3630	2.2112
UMSN [22]	27.0050	0.8276	0.1460	39.4150	1.4908
UMSN*	30.5884	0.9140	0.0833	16.8517	1.2716
MSPL-GAN [23]	28.2286	0.8936	0.1092	27.9441	1.2947
MSPL-GAN*	34.2711	<u>0.9359</u>	<u>0.0638</u>	<u>10.3597</u>	<u>1.0983</u>
CFMD-GAN ₁₂₈	<u>33.8475</u>	0.9379	0.04910	6.5449	1.0246

Continuous Facial Motion Deblurring

Tae Bok Lee, Suhy Han, and Yong Seok Heo
IEEE Access, 2022

Experiments

Comparison with Single-to-Single Generic Deblurring Methods



Input

DMPHN

MIMO

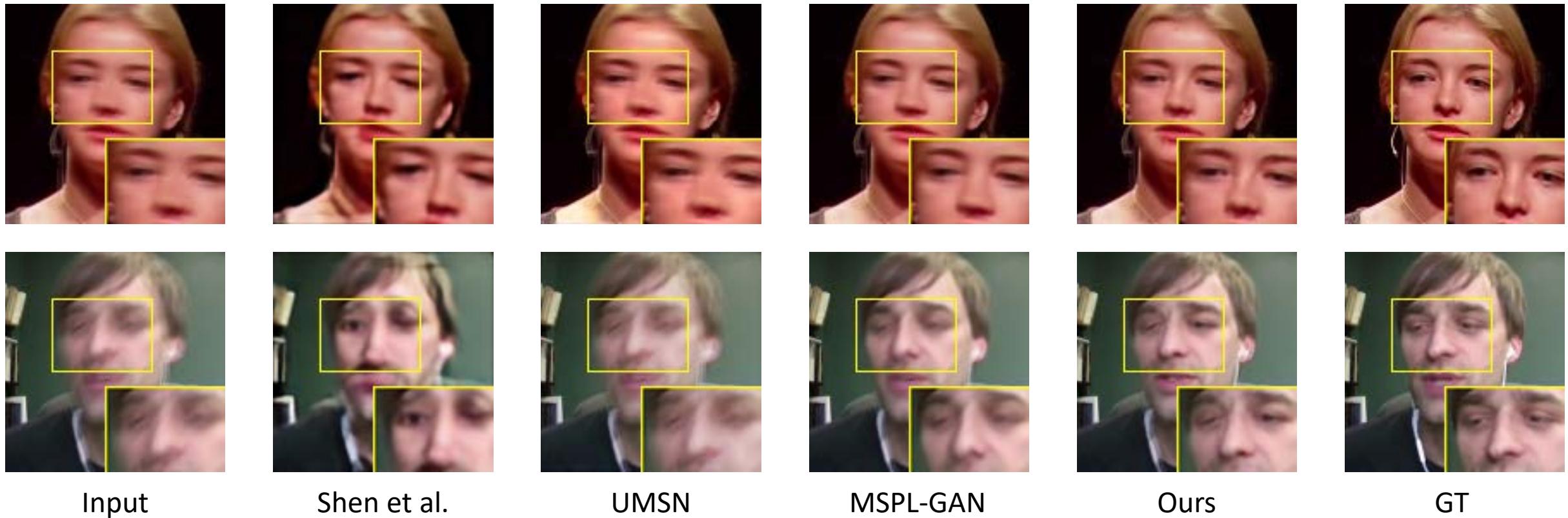
Ours

GT

Continuous Facial Motion Deblurring

Tae Bok Lee, Suhy Han, and Yong Seok Heo
IEEE Access, 2022

Experiments

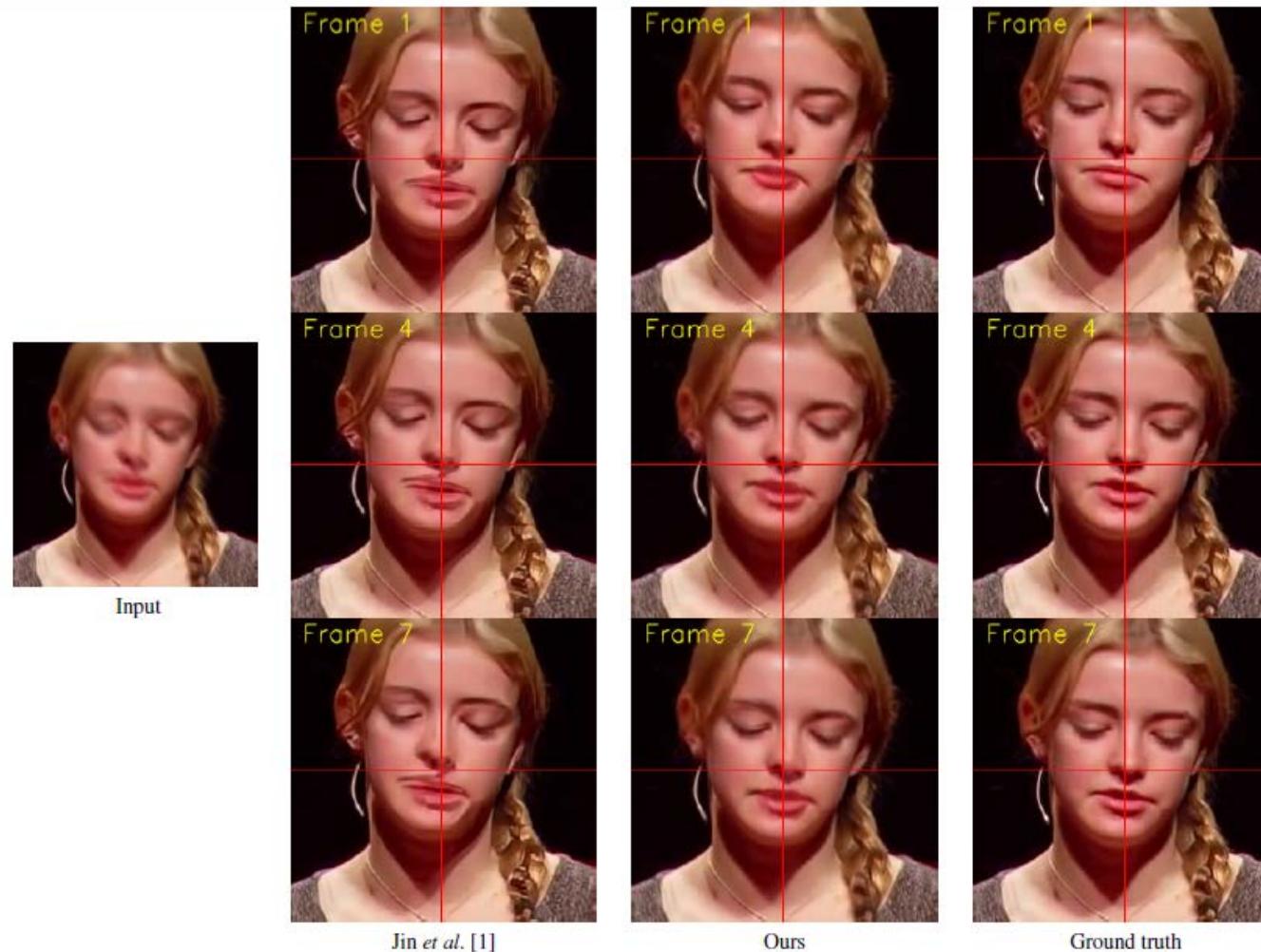


Continuous Facial Motion Deblurring

Tae Bok Lee, Suhy Han, and Yong Seok Heo
IEEE Access, 2022

Experiments

Comparison with Single-to-Video Generic Deblurring Methods



Continuous Facial Motion Deblurring

Tae Bok Lee, Suhy Han, and Yong Seok Heo
IEEE Access, 2022

Experiments

Comparison with Single-to-Video Generic Deblurring Methods

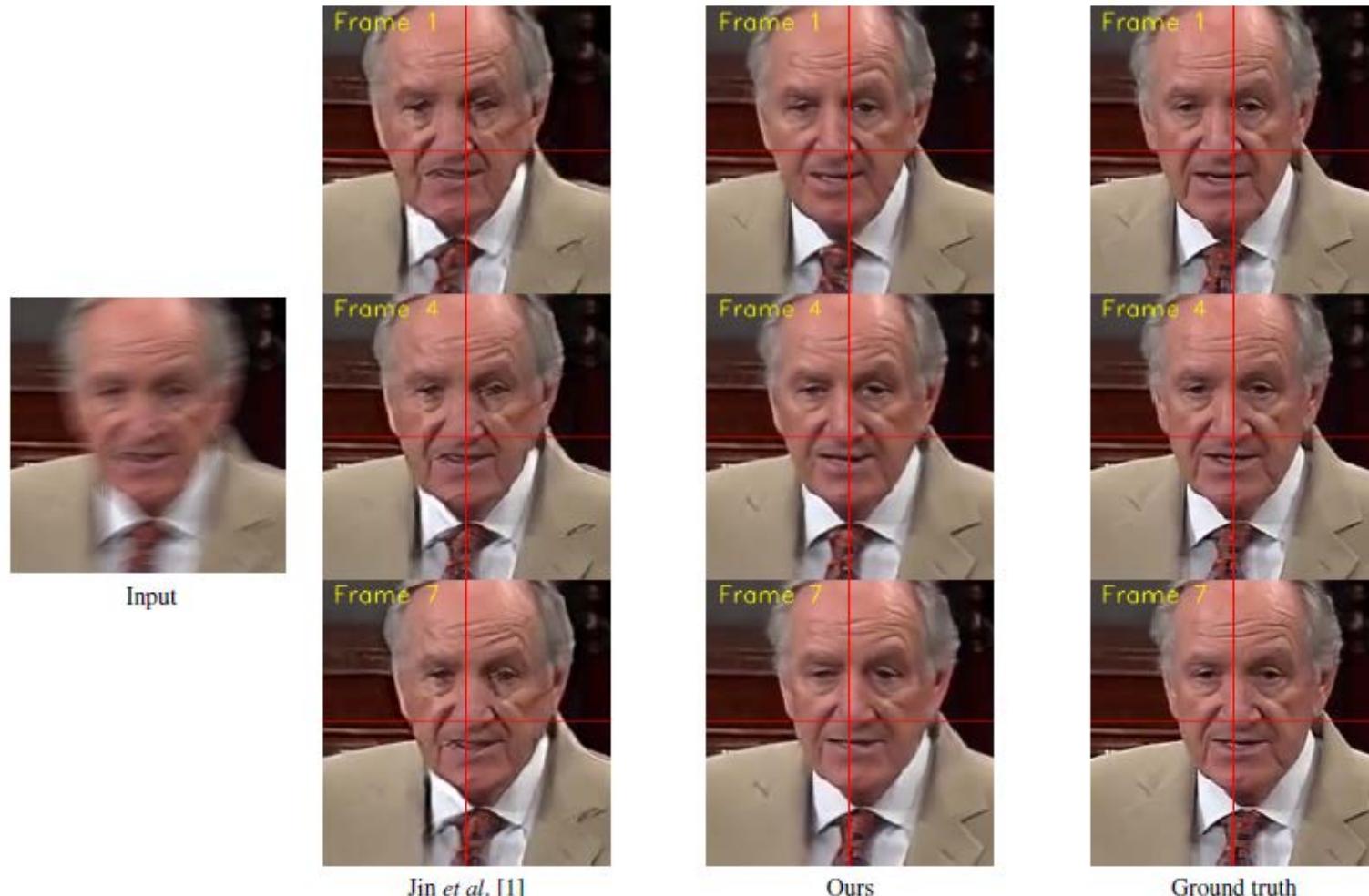


Continuous Facial Motion Deblurring

Tae Bok Lee, Suhy Han, and Yong Seok Heo
IEEE Access, 2022

Experiments

Comparison with Single-to-Video Generic Deblurring Methods

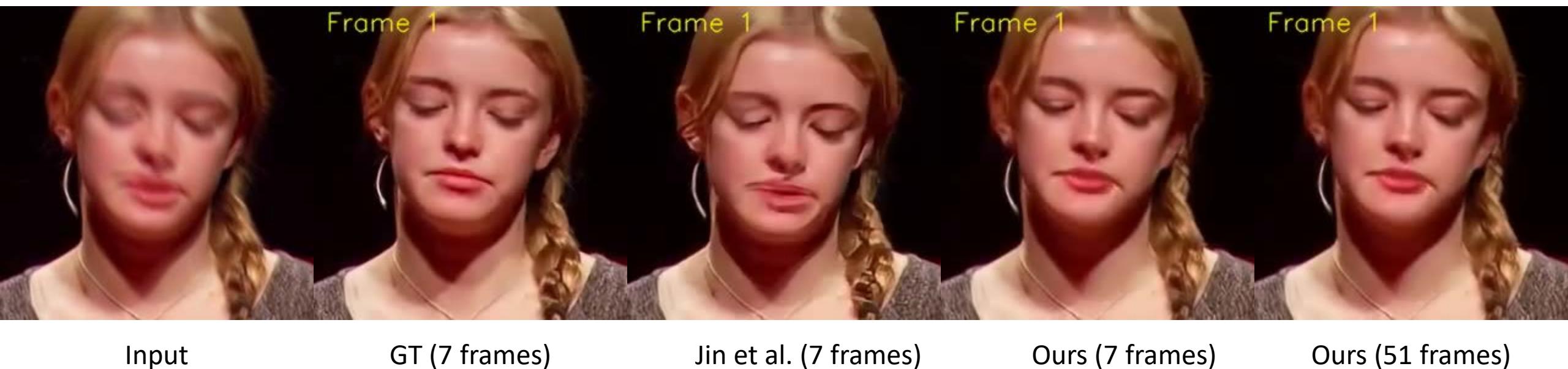


Continuous Facial Motion Deblurring

Tae Bok Lee, Suhy Han, and Yong Seok Heo
IEEE Access, 2022

Experiments

Comparison with Single-to-Video Generic Deblurring Methods

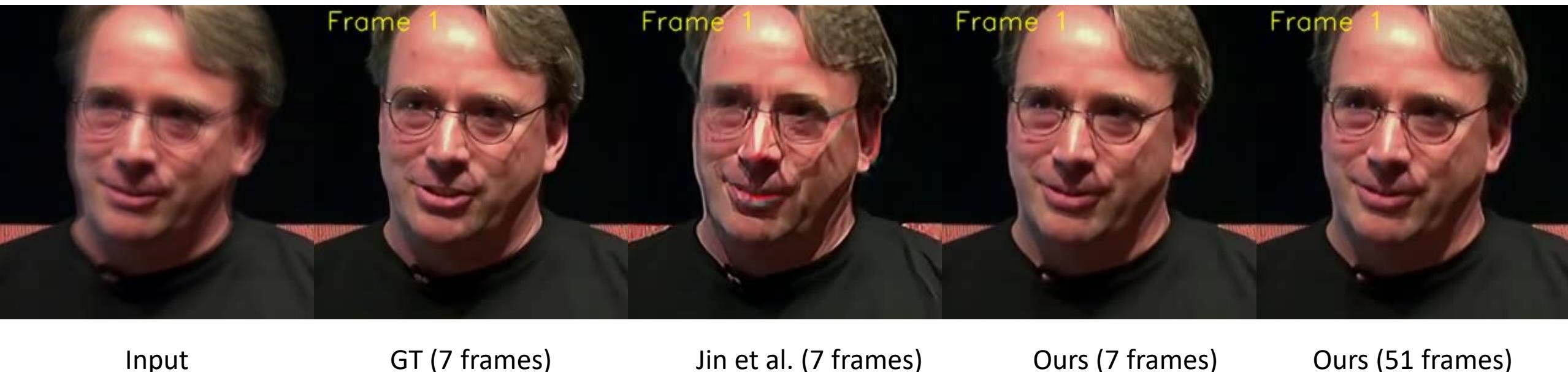


Continuous Facial Motion Deblurring

Tae Bok Lee, Suhy Han, and Yong Seok Heo
IEEE Access, 2022

Experiments

Comparison with Single-to-Video Generic Deblurring Methods



Continuous Facial Motion Deblurring

Tae Bok Lee, Suhy Han, and Yong Seok Heo
IEEE Access, 2022

Experiments

Comparison with Single-to-Video Generic Deblurring Methods



Continuous Facial Motion Deblurring

Tae Bok Lee, Suhy Han, and Yong Seok Heo
IEEE Access, 2022

Experiments

Comparison with Single-to-Video Generic Deblurring Methods



Input

GT (7 frames)

Jin et al. (7 frames)

Ours (7 frames)

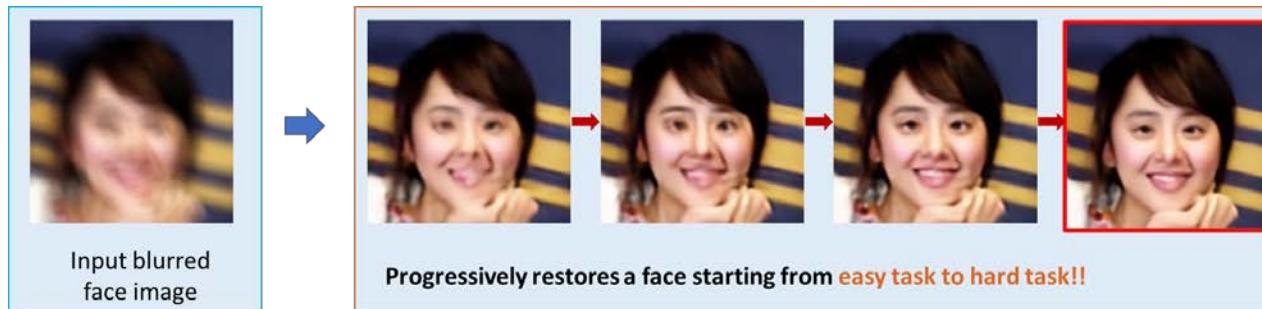
Ours (51 frames)

Progressive Semantic Face Deblurring

Tae Bok Lee, Soo Hyun Jung, and Yong Seok Heo
IEEE Access, 2020

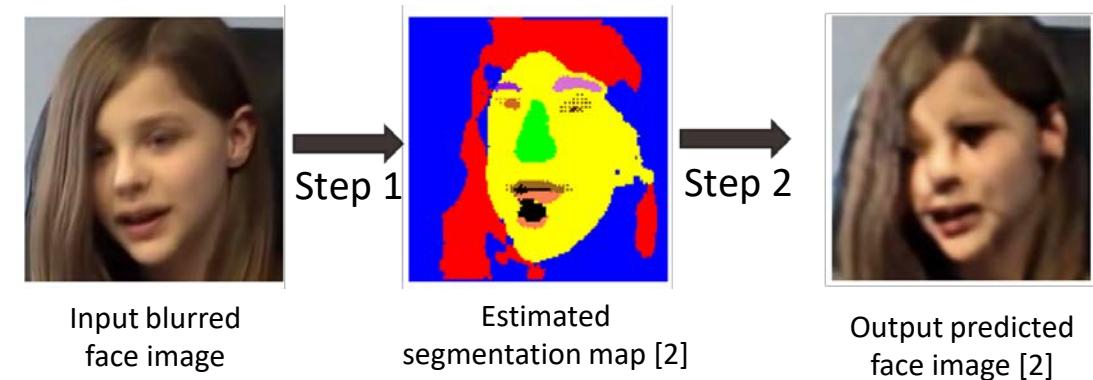
Single Face Image Deblurring

- Restore a single sharp face image from a blurred one.



Previous Works

- Leverage prior knowledge of the face
 - Reference face, landmark, 3D face models, segmentation map
- Two-step framework
 - Step 1. Estimate useful prior information
⇒ Difficult to estimate accurate prior (e.g. segmentation map) from blurred images
 - Step 2. Restore the sharp image
⇒ Inaccurate prior negatively affects to the restoration performance



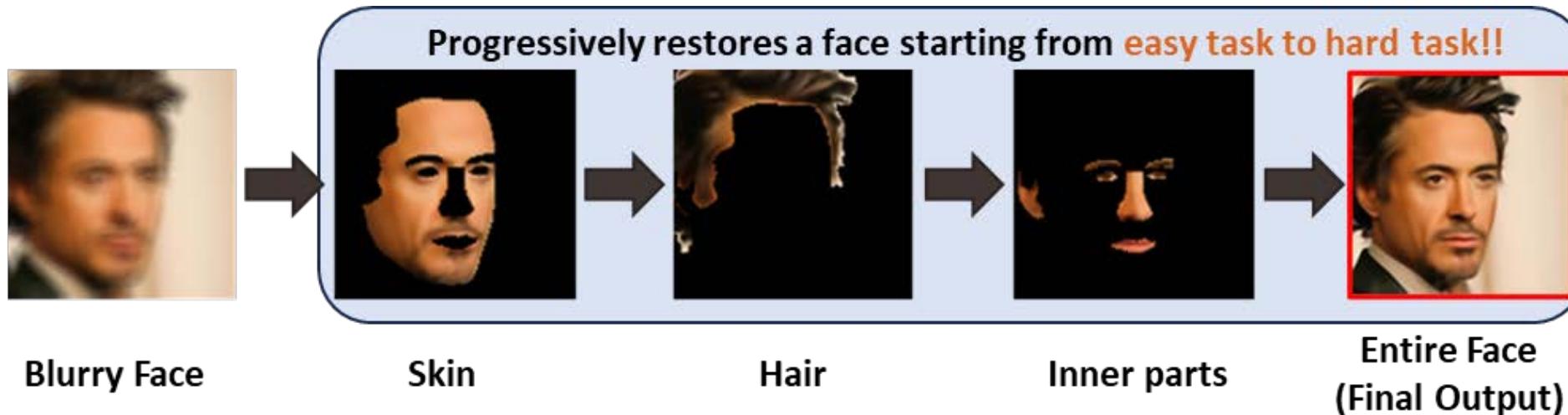
[1] W. Ren J. Yang S. Deng D. Wipf X. Cao and X. Tong "Face video deblurring using 3D facial priors" Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV) pp. 9388-9397 Oct. 2019.

[2] Z. Shen W.-S. Lai T. Xu J. Kautz and M.-H. Yang "Deep semantic face deblurring" Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. pp. 8260-8269 Jun. 2018.

Progressive Semantic Face Deblurring

Tae Bok Lee, Soo Hyun Jung, and Yong Seok Heo
IEEE Access, 2020

Proposed Progressive Semantic Face Deblurring

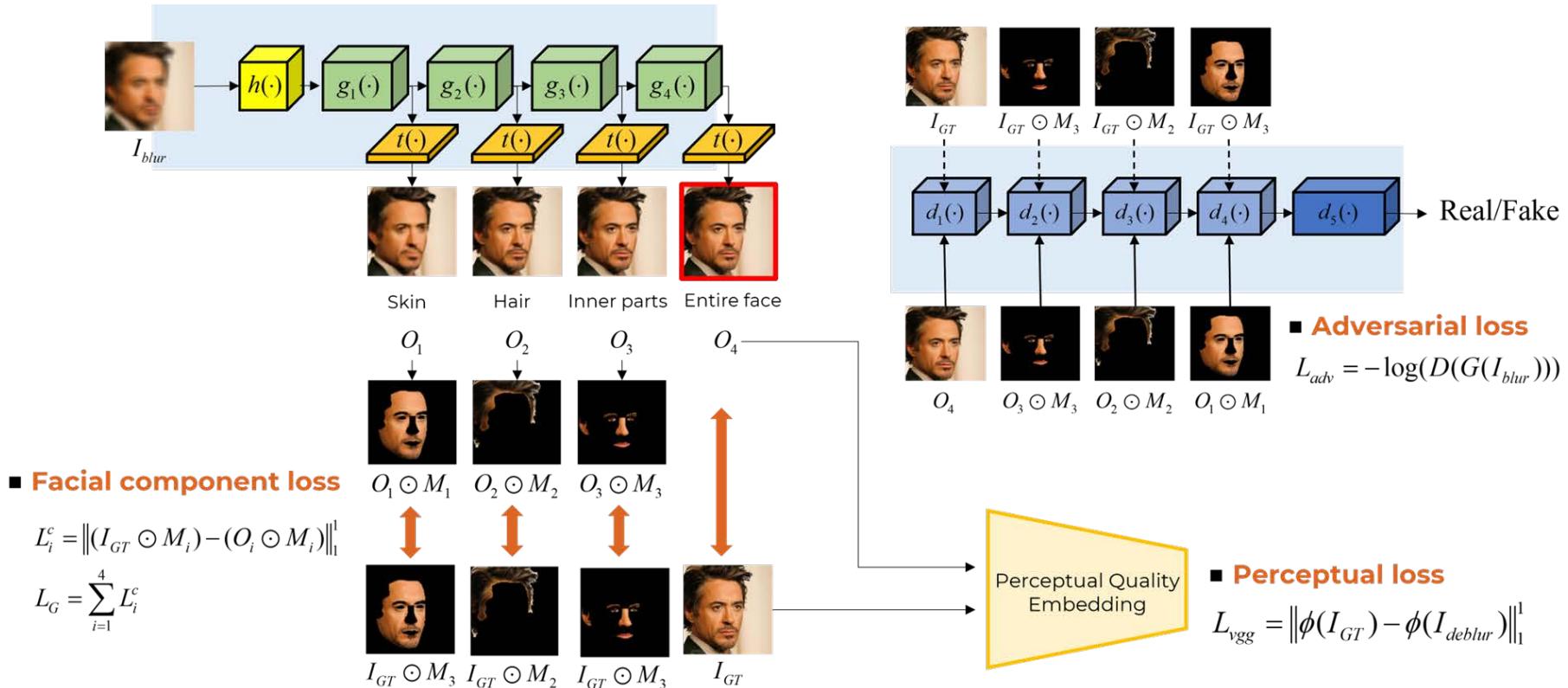


- Leverage semantic priors **without performing the segmentation process**
 - Prevent the side effects of an inaccurate segmentation map
- Progressively restore the face **from easy component (low frequency texture) to hard component (high frequency and small texture)**
 - Better understand the underlying semantic structure of the face
- [\[Paper\]](#) [\[Github\]](#)

Progressive Semantic Face Deblurring

Tae Bok Lee, Soo Hyun Jung, and Yong Seok Heo
IEEE Access, 2020

Overall Framework of Proposed MSPL-GAN



Semantic Progressive Generator

- Without prior estimation step
- Incrementally restore the face in semantic order
 - Starting with an **easy task** and **gradually refining the details**
 - Skin → Hair → Inner parts → Entire face

Multi-Semantic Discriminator

- Classify whether facial components are real or fake
- Provide adversarial feedback to generator for restoration of realistic facial components

Progressive Semantic Face Deblurring

Tae Bok Lee, Soo Hyun Jung, and Yong Seok Heo

IEEE Access, 2020

Experiments

Comparison on Shen Testset

Method	Helen				CelebA			
	PSNR(↑)	SSIM(↑)	$d_{VGG}(\downarrow)$	LPIPS(↓)	PSNR(↑)	SSIM(↑)	$d_{VGG}(\downarrow)$	LPIPS(↓)
Shen <i>et al.</i> [7]	25.58	0.861	91.06	0.1527	24.34	0.860	117.50	0.1832
Lu <i>et al.</i> [68]	20.25	0.705	241.93	0.1654	19.96	0.742	305.96	0.1688
Xia <i>et al.</i> [58]	26.13	0.886	55.97	0.1052	25.18	0.892	68.05	0.1199
Yasarla <i>et al.</i> [8]	27.75	0.897	86.87	0.1086	26.62	0.908	66.33	0.1401
MSPL_GAN	25.91	0.881	47.80	0.0828	24.91	0.885	57.54	0.0962

Network Parameters & Inference Time

Method	Implementation	Inference time (S)	Parameters (M)
Shen <i>et al.</i> [7]	MATLAB(GPU)	0.05	14.8
Lu <i>et al.</i> [68]	Pytorch(GPU)	0.02	53.0
Xia <i>et al.</i> [58]	Tensorflow(GPU)	0.19	41.8
Yasarla <i>et al.</i> [8]	Pytorch(GPU)	0.16	14.4
Ours	Pytorch(GPU)	0.08	18.5

Comparison on MSPL Testset

Method	MSPL-Center								FFHQ			
	CelebA				CelebA-HQ				FFHQ			
PSNR(↑)	SSIM(↑)	$d_{VGG}(\downarrow)$	LPIPS(↓)	PSNR(↑)	SSIM(↑)	$d_{VGG}(\downarrow)$	LPIPS(↓)	PSNR(↑)	SSIM(↑)	$d_{VGG}(\downarrow)$	LPIPS(↓)	
Shen <i>et al.</i> [7]	19.75	0.740	113.66	0.3007	19.95	0.755	267.41	0.2865	19.57	0.723	220.87	0.3417
Lu <i>et al.</i> [68]	17.93	0.617	123.35	0.2284	18.63	0.649	243.06	0.1902	18.26	0.630	177.00	0.2256
Xia <i>et al.</i> [58]	25.03	0.873	39.58	0.1790	25.79	0.886	83.46	0.1608	24.66	0.859	57.66	0.2081
Yasarla <i>et al.</i> [8]	22.73	0.817	55.01	0.2132	23.02	0.827	102.97	0.1956	22.19	0.795	86.43	0.2506
MSPL_GAN	28.07	0.921	18.19	0.1152	28.82	0.929	40.93	0.0968	27.36	0.908	25.39	0.1325

Method	MSPL-Random								FFHQ			
	CelebA				CelebA-HQ				FFHQ			
PSNR(↑)	SSIM(↑)	$d_{VGG}(\downarrow)$	LPIPS(↓)	PSNR(↑)	SSIM(↑)	$d_{VGG}(\downarrow)$	LPIPS(↓)	PSNR(↑)	SSIM(↑)	$d_{VGG}(\downarrow)$	LPIPS(↓)	
Shen <i>et al.</i> [7]	18.89	0.711	90.37	0.3310	19.18	0.729	157.49	0.3185	19.03	0.713	127.71	0.3356
Lu <i>et al.</i> [68]	17.41	0.631	46.05	0.2693	18.04	0.664	72.56	0.2297	17.94	0.654	65.06	0.2589
Xia <i>et al.</i> [58]	23.66	0.849	30.94	0.2044	24.48	0.861	60.95	0.1940	23.95	0.855	44.62	0.2016
Yasarla <i>et al.</i> [8]	21.24	0.777	45.05	0.2448	21.46	0.789	72.56	0.2296	21.28	0.778	65.06	0.2407
MSPL_GAN	28.95	0.936	11.41	0.1090	29.80	0.945	26.91	0.0938	29.22	0.941	15.44	0.0988

Progressive Semantic Face Deblurring

Tae Bok Lee, Soo Hyun Jung, and Yong Seok Heo
IEEE Access, 2020

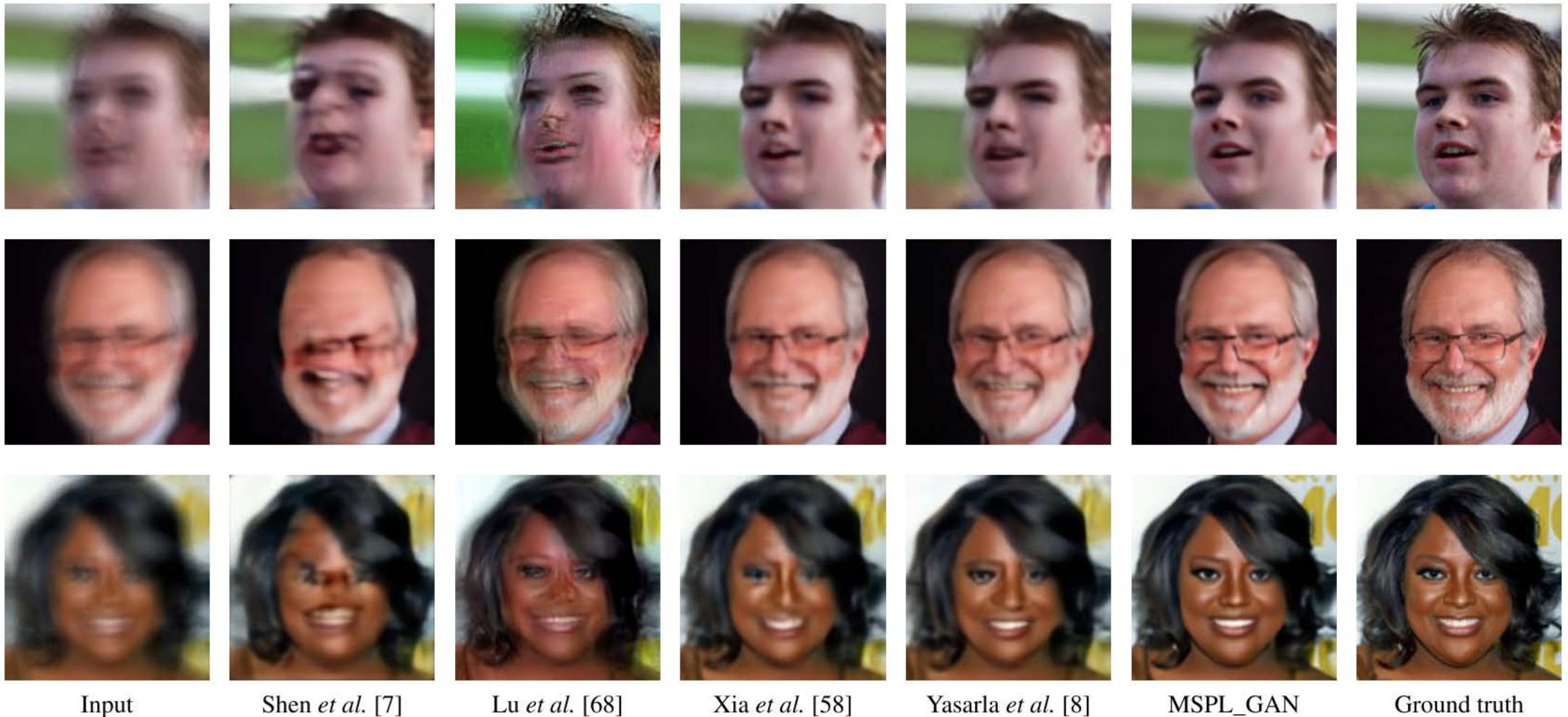
Comparison on Shen Testset



Progressive Semantic Face Deblurring

Tae Bok Lee, Soo Hyun Jung, and Yong Seok Heo
IEEE Access, 2020

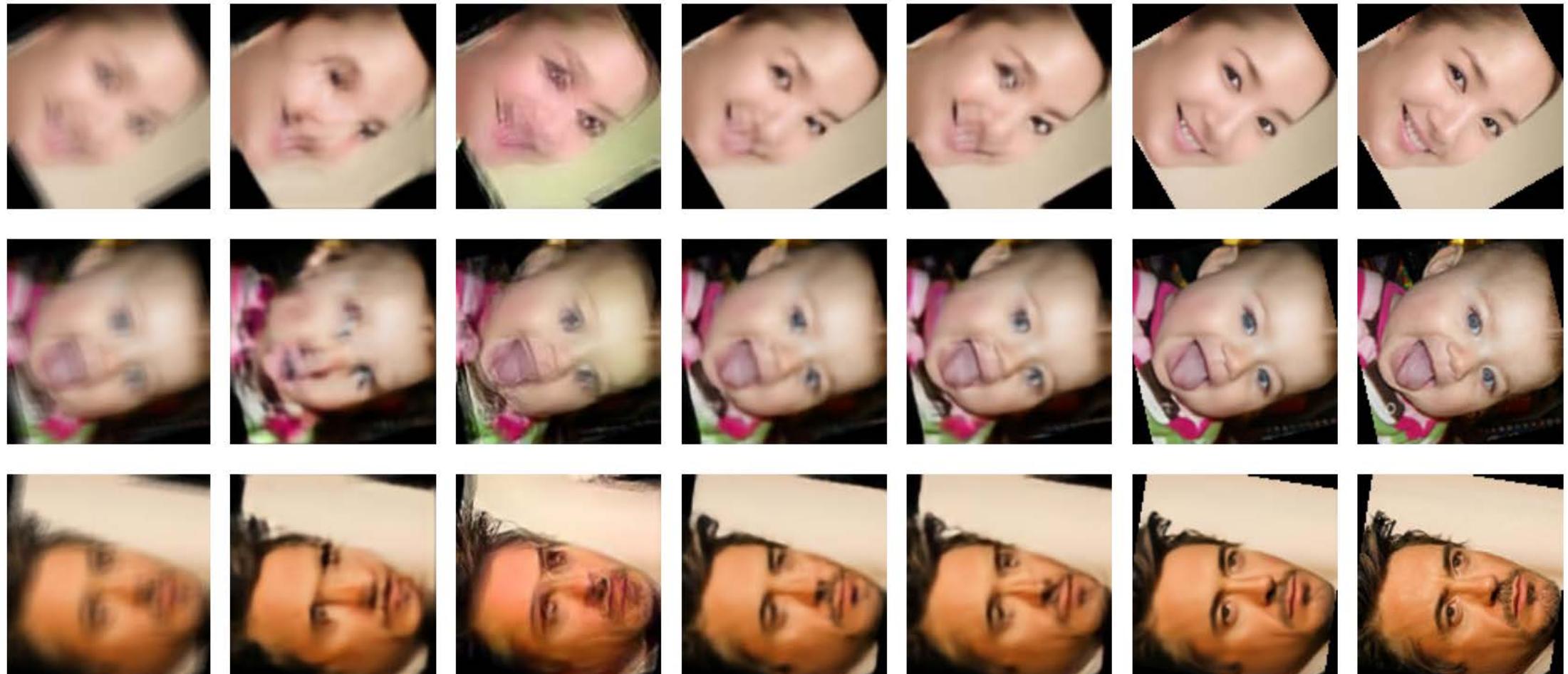
Comparison on MSPL Center Testset



Progressive Semantic Face Deblurring

Tae Bok Lee, Soo Hyun Jung, and Yong Seok Heo
IEEE Access, 2020

Comparison on MSPL Random Testset



Input

Shen *et al.* [7]

Lu *et al.* [68]

Xia *et al.* [58]

Yasarla *et al.* [8]

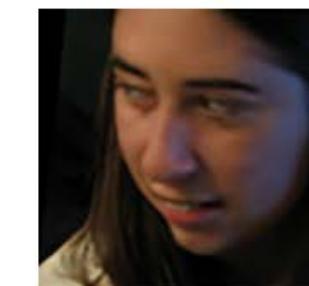
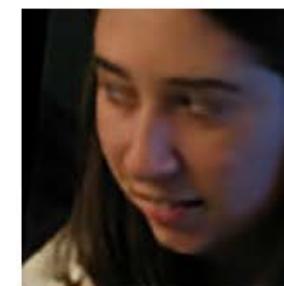
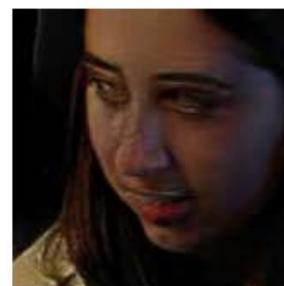
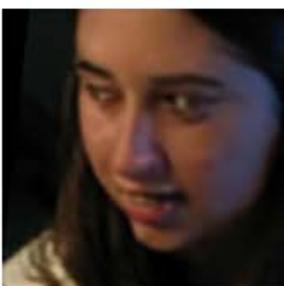
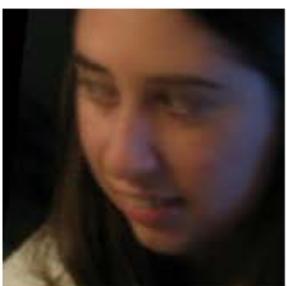
MSPL_GAN

Ground truth

Progressive Semantic Face Deblurring

Tae Bok Lee, Soo Hyun Jung, and Yong Seok Heo
IEEE Access, 2020

Comparison on Realworld Blur Testset



Input

Shen *et al.* [7]

Lu *et al.* [68]

Xia *et al.* [58]

Yasarla *et al.* [8]

MSPL_GAN

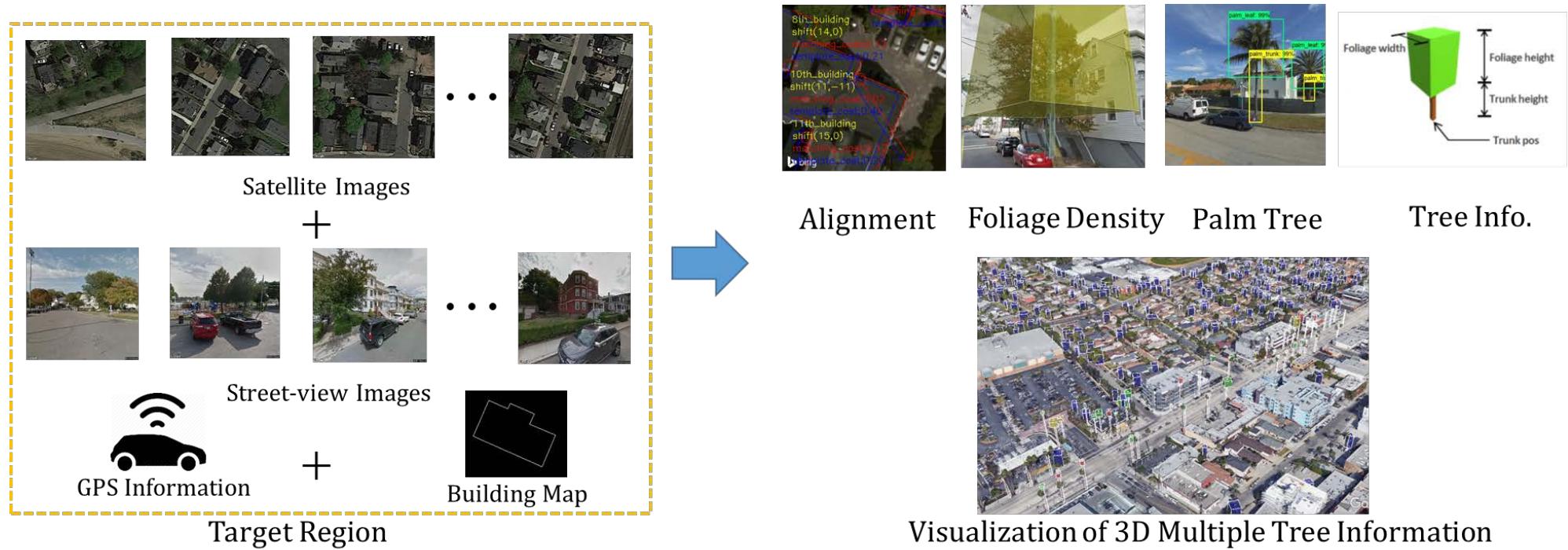
Participated Projects

딥러닝 기반 정밀 Semantic Map 추출 엔진 개발

삼성 전자, 2018-2019.

Project Goal

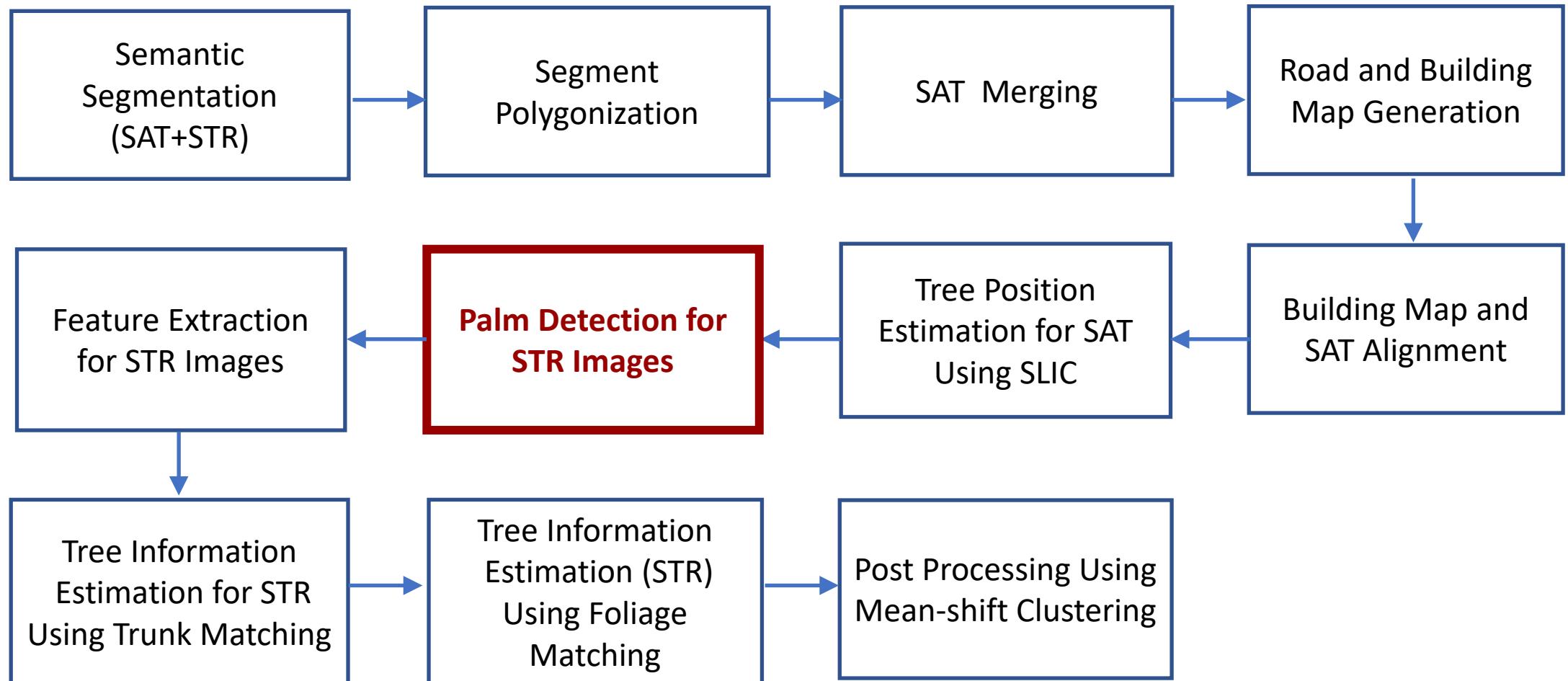
- Alignment between Building Maps and Aerial Images
- Estimation of Multiple Information of Trees
 - Tree Position, Foliage Width, Foliage Height, Trunk Width, Trunk Height
 - Palm Tree, Foliage Density



딥러닝 기반 정밀 Semantic Map 추출 엔진 개발

삼성 전자, 2018-2019.

Overview



SAT : Satellite images

STR : Street-view images

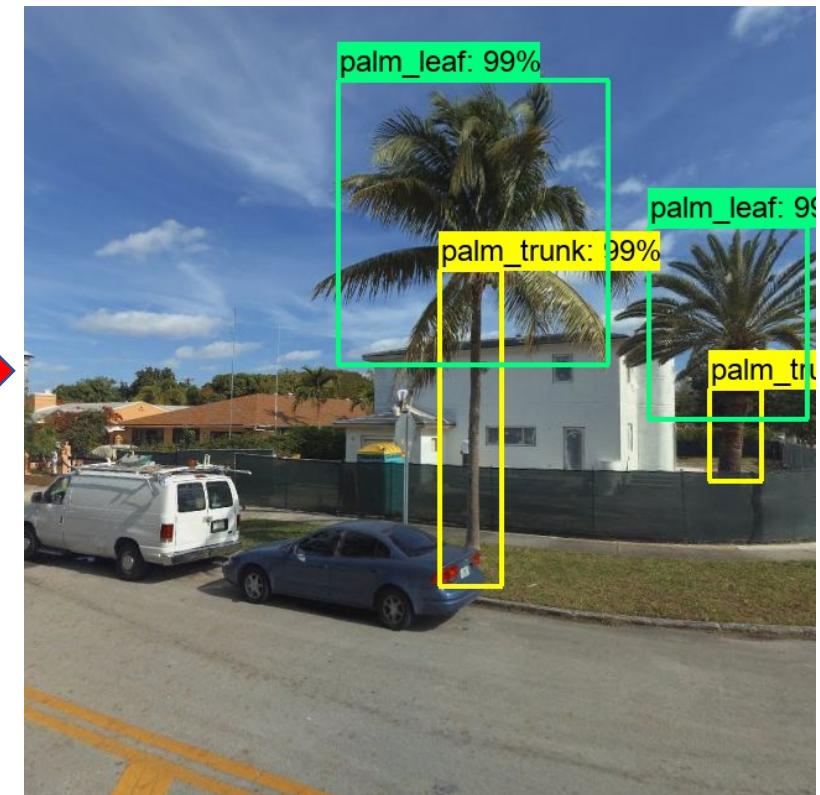
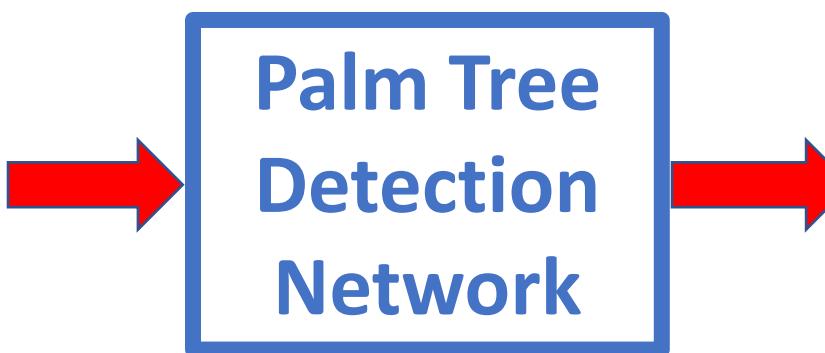
SLIC: R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC Superpixels Compared to State-of-the-art Superpixel Methods," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, num. 11, p. 2274 - 2282, May 2012.

딥러닝 기반 정밀 Semantic Map 추출 엔진 개발

삼성 전자, 2018-2019.

Palm Tree Detection for Street-View Images

- Goal: Detect Palm Leaf and Trunk



Original Image

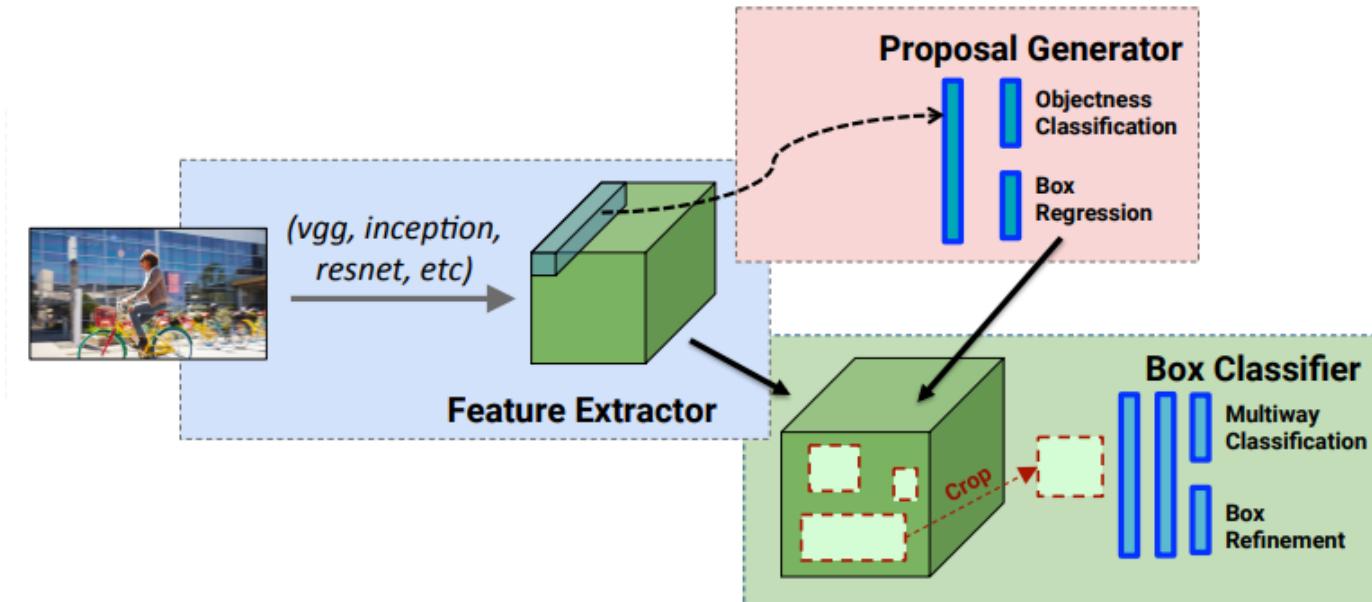
Detected Image

딥러닝 기반 정밀 Semantic Map 추출 엔진 개발

삼성 전자, 2018-2019.

Configuration

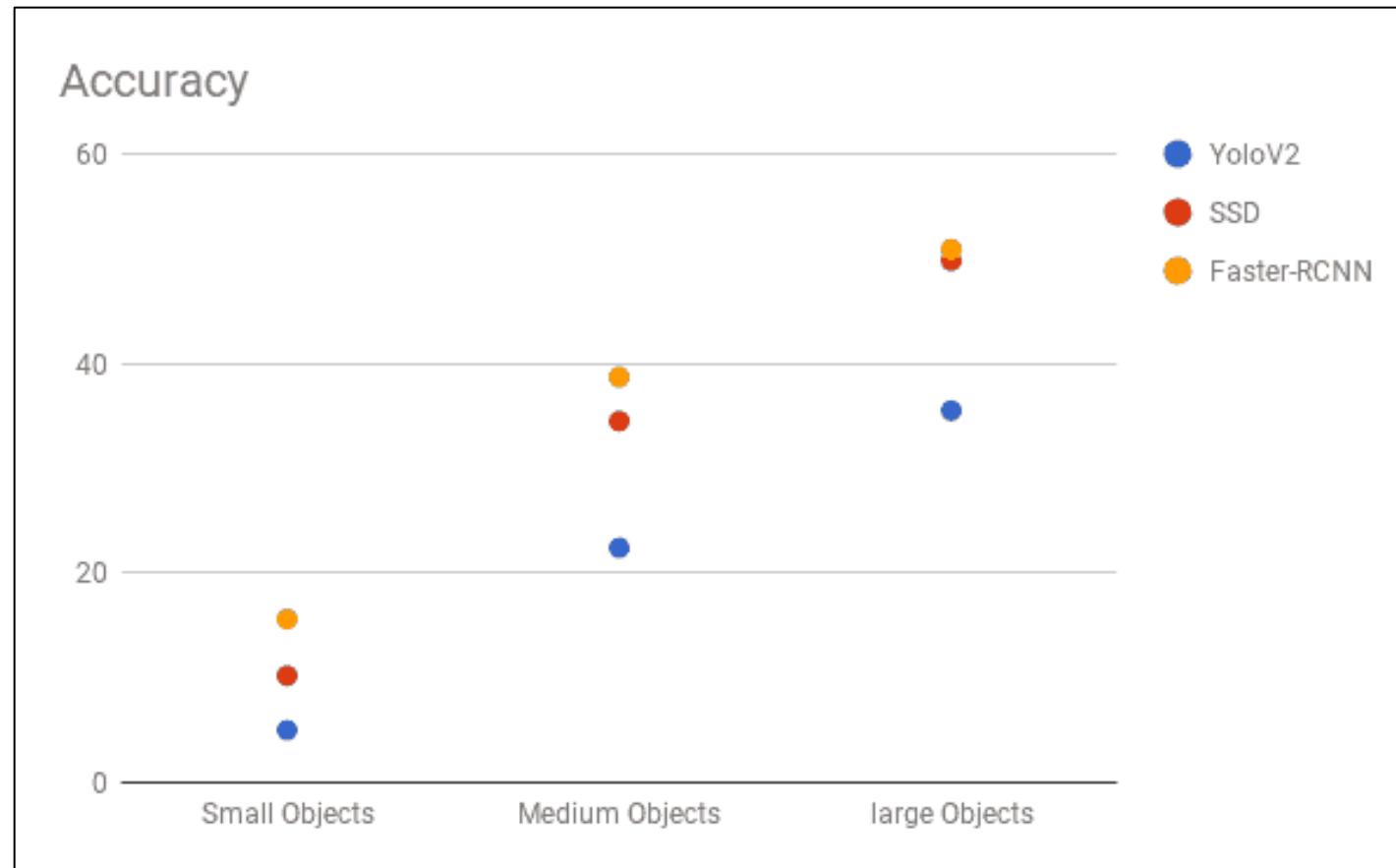
- Architecture : Faster R-CNN*
- Feature Extractor Network : Inception V2, Resnet101
- 3 Classes : Palm leaf, Palm trunk, other objects



*S.Ren, K.He, R.Grishick, J.Sun, "Towards real-time object detection with region proposed networks," ICCV 2015.

딥러닝 기반 정밀 Semantic Map 추출 엔진 개발 삼성 전자, 2018-2019.

General Performance - Architecture



Comparison of YoloV2, SSD, Faster-RCNN

딥러닝 기반 정밀 Semantic Map 추출 엔진 개발

삼성 전자, 2018-2019.

Dataset Configuration

Palm Tree Dataset	Training Set	Test Set	Total
Palm Image	569	60	629
Negative Image	100	60	160
Total	669	120	779

Table1. Dataset Configuration

- Total Training Images : 659
 - Bing Palm Tree images : 569
 - Bing Negative object images: 100
- Total Test Images : 120
 - Bing Palm Tree images: 60
 - Bing Negative object images : 60

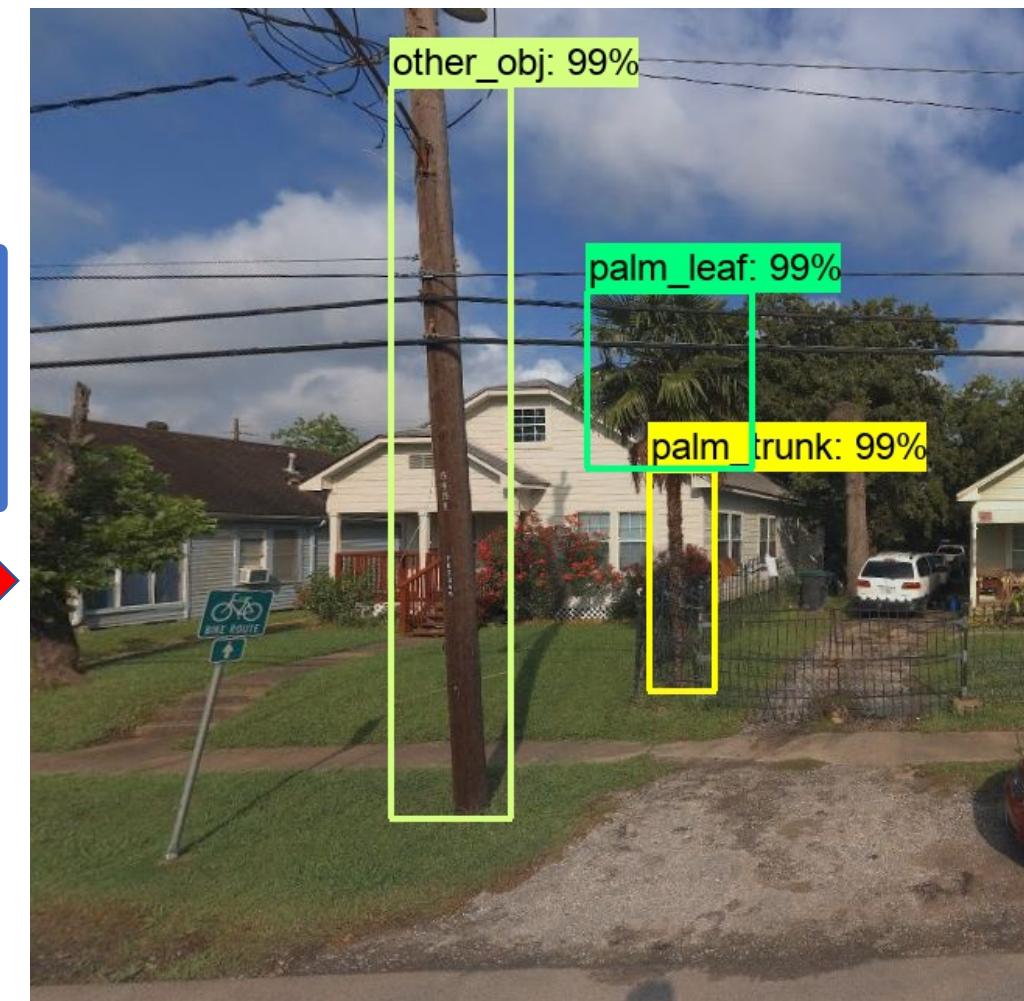
딥러닝 기반 정밀 Semantic Map 추출 엔진 개발

삼성 전자, 2018-2019.

Test Results Images



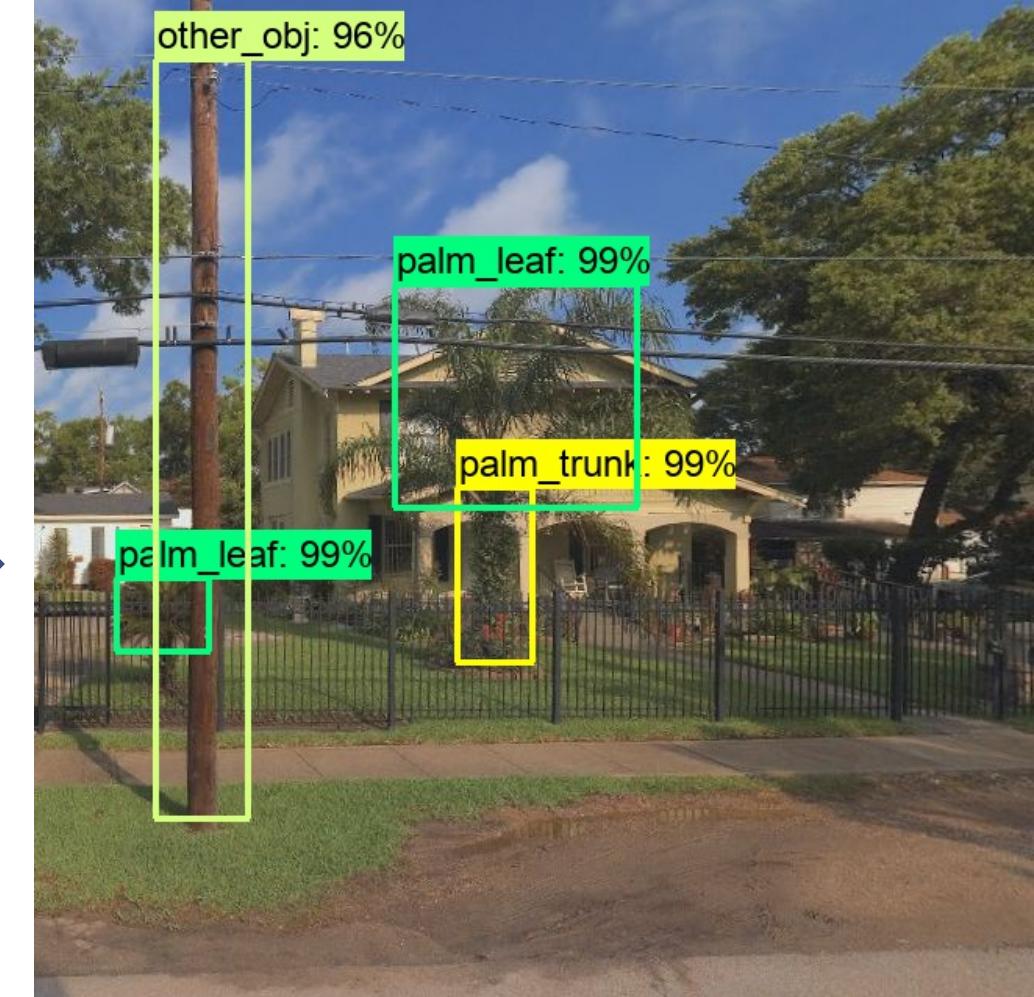
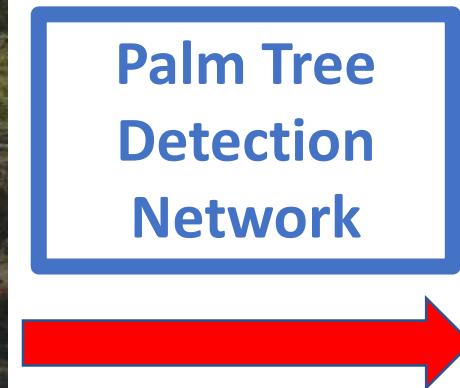
Palm Tree
Detection
Network



딥러닝 기반 정밀 Semantic Map 추출 엔진 개발

삼성 전자, 2018-2019.

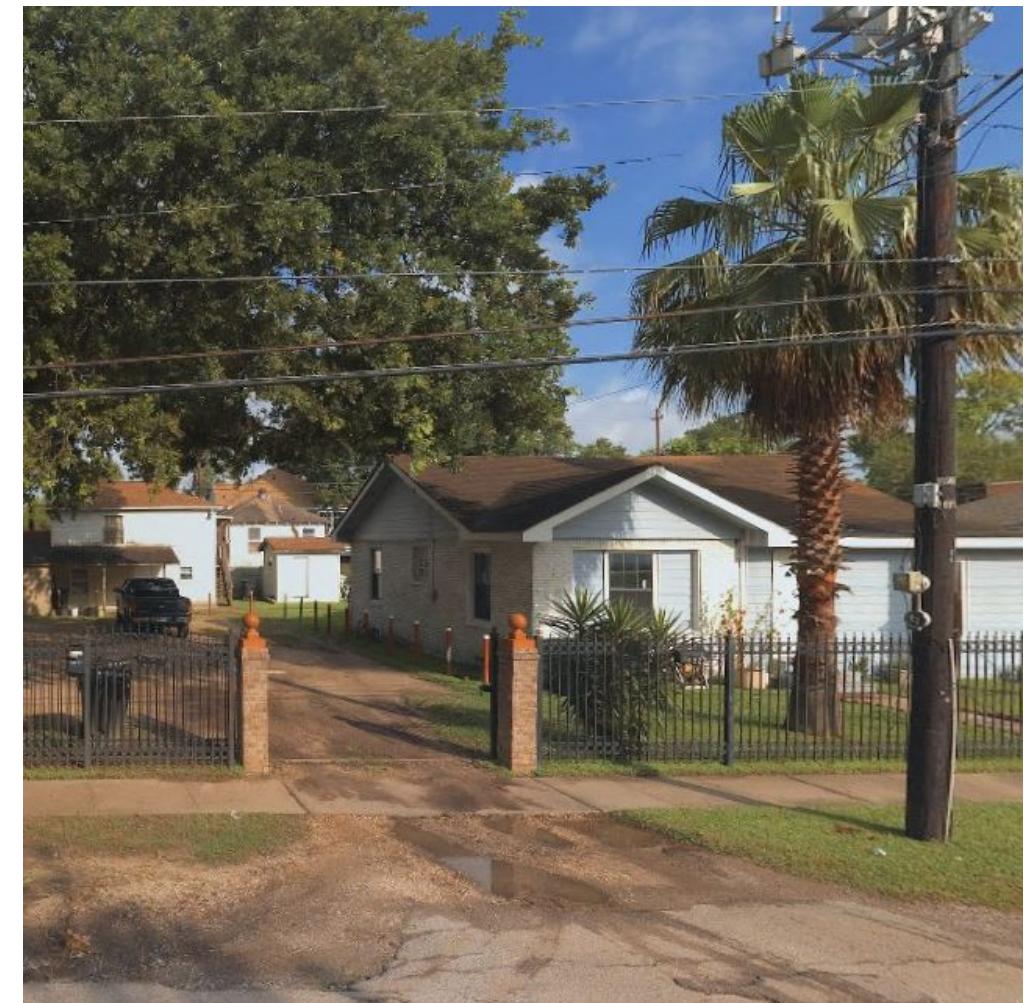
Test Results Images



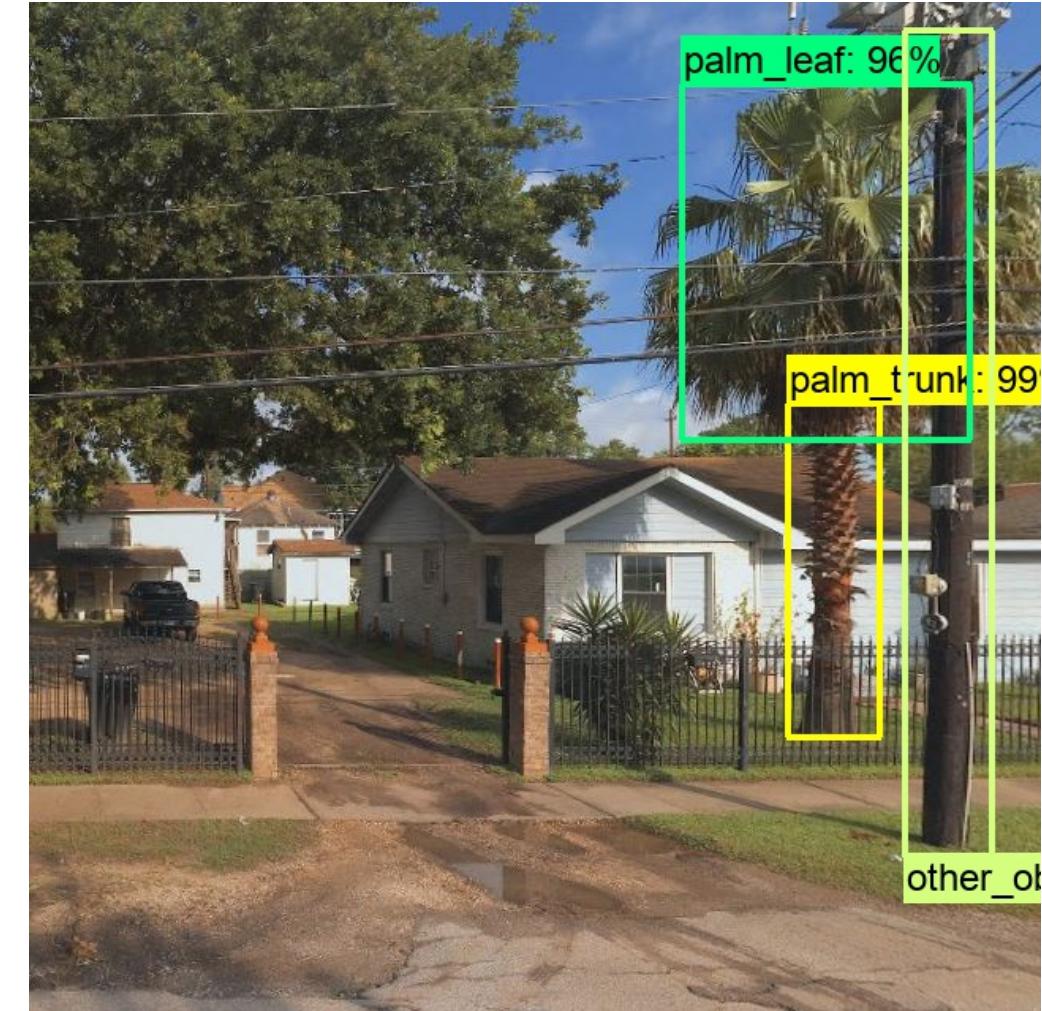
딥러닝 기반 정밀 Semantic Map 추출 엔진 개발

삼성 전자, 2018-2019.

Test Results Images



Palm Tree
Detection
Network



UHD 방송 환경에서 콘텐츠에 대한 시청자의 반응 및 의도 기반 미디어 인터랙션 기술 과학기술정보통신부, 2018-2020.

Project Goal

- UHD 방송 환경에서 콘텐츠에 대한 시청자의 반응과 의도를 파악하여 개인별 맞춤형 서비스를 제공하는 "지능형 미디어 인터랙션" 기술



UHD 방송 환경에서 콘텐츠에 대한 시청자의 반응 및 의도 기반 미디어 인터랙션 기술 과학기술정보통신부, 2018-2020.

연차별 연구 내용

항목	연구내용		
	1차년도	2차년도	3차년도
영상 기반의 방송 콘텐츠 시청자 반응 및 의도 인지 기술	<ul style="list-style-type: none">영상 기반의 실시간 얼굴 특징점 검출 기능 구현<ul style="list-style-type: none">Random Forests 기반의 실시간 얼굴 특징점 검출정서/감정 인식을 위한 데이터베이스 구축<ul style="list-style-type: none">다양한 감정적 반응을 포함하는 동영상 수집Valence/Arousal Labeling	<ul style="list-style-type: none">얼굴 영상으로부터의 정서/감정 인식 기술 개발<ul style="list-style-type: none">특징점을 활용한 MLP 기반의 정서/감정 인식영상을 활용한 CNN 기반의 정서/감정 인식시선 방향 추정을 위한 데이터베이스 구축<ul style="list-style-type: none">모바일, 태블릿, TV 등 다양한 환경에서의 사용자 얼굴 영상 및 시선 데이터 수집	<ul style="list-style-type: none">얼굴 영상으로부터의 시선 방향 추정 기술 개발<ul style="list-style-type: none">특징점을 활용한 MLP 기반의 시선 방향 추정영상을 활용한 CNN 기반의 시선 방향 추정영상 기반의 실시간 정서/감정 및 관심 영역 검출<ul style="list-style-type: none">다중 사용자 얼굴 영상으로부터의 실시간 정서/감정 및 시선 방향 추정
생체신호 기반의 방송 콘텐츠 시청자 반응 및 의도 인지 기술	<ul style="list-style-type: none">UHD 방송 콘텐츠 시청자의 생체신호 데이터베이스 구축 및 정서/감정 특징 추출 기술 개발<ul style="list-style-type: none">시청자 반응 및 의도 추출을 위한 멀티모달 생체신호 수집 및 데이터베이스 구축시청자 생체정보 기반의 정서/감정 반응 및 의도 추출 지표 개발	<ul style="list-style-type: none">멀티모달 생체신호 기반 시청자 반응 및 의도 인식 기술 개발<ul style="list-style-type: none">UHD 방송 시청자의 명시적 및 묵시적 의도 인식 모델 개발멀티모달 생체신호 특징 융합 기반 감정 및 의도 분류 모델 개발멀티모달 생체신호 DB를 이용한 RNN 기반 의도 인식 고도화	<ul style="list-style-type: none">생체신호 기반 시청자 맞춤형 정서/감정 인식 최적화<ul style="list-style-type: none">멀티모달 생체신호를 이용한 시청자 맞춤형 감정 및 의도 인식 최적화 모델 개발시청자 맞춤형 UHD 방송 콘텐츠 인터랙션 예측 모델 개발<ul style="list-style-type: none">시청자 감정 및 의도를 고려한 인터랙션 예측 모델 설계 및 검증
방송 콘텐츠 영상 내 시청자 관심 대상 검출 및 인식 기술	<ul style="list-style-type: none">단일 물체 분류를 위한 영상 기반의 특징점 추출<ul style="list-style-type: none">TV 영상 컨텐츠에 특화된 데이터베이스 구축딥러닝 기반의 단일 물체에 대한 특징점 추출 네트워크 개발	<ul style="list-style-type: none">특징점 기반의 물체 분류 기술 개발<ul style="list-style-type: none">학습된 특징점 기반 물체 클래스 분류 기술 개발분류된 물체의 정확한 위치 추정 기술 개발<ul style="list-style-type: none">분류된 물체의 위치에 대한 Bounding Box 추정	<ul style="list-style-type: none">영상 기반의 다중 물체 검출 및 위치 예측 기술 개발<ul style="list-style-type: none">다중 물체의 위치와 클래스를 종합하여 예측실시간 다중 물체 검출 알고리즘 및 시스템 최적화<ul style="list-style-type: none">실시간 동작을 위한 알고리즘 및 코드 최적화 작업
반응 및 의도 기반 미디어 서비스 엔진 및 플랫폼 개발	<ul style="list-style-type: none">시청자 반응 및 의도 기반 미디어 서비스 엔진 및 플랫폼 설계<ul style="list-style-type: none">시청자 프로파일 및 컨텍스트 구조와 처리 엔진 설계반응 및 의도 기반 미디어 서비스 개발을 위한 테스트 플랫폼 구축	<ul style="list-style-type: none">시청자 반응 및 의도 기반 미디어 서비스 엔진 및 시나리오 구현<ul style="list-style-type: none">미디어 서비스를 위한 시청자 프로파일 및 컨텍스트 처리 엔진 구현시청자 반응 및 의도 기반의 인터랙티브 미디어 서비스 시나리오 설계 및 구현	<ul style="list-style-type: none">인터랙티브 방송서비스 연동이 가능한 테스트 플랫폼 및 BM 개발<ul style="list-style-type: none">인터랙티브 방송 서비스와의 연동이 가능한 테스트 플랫폼 구축인터랙티브 방송 환경에서의 시청자 반응 및 의도 기반 BM 개발 및 검증

UHD 방송 환경에서 콘텐츠에 대한 시청자의 반응 및 의도 기반 미디어 인터랙션 기술

과학기술정보통신부, 2018-2020.

방송 콘텐츠 분석 및 인식

▪ 전체 연구 개요

<연구목표>

- 방송 콘텐츠 영상에서 메타데이터를 활용한 실시간 다중 관심 물체 분류 및 위치 추정

<연구내용>

1. 단일 물체 분류를 위한 영상 기반의 특징 추출 기술 연구

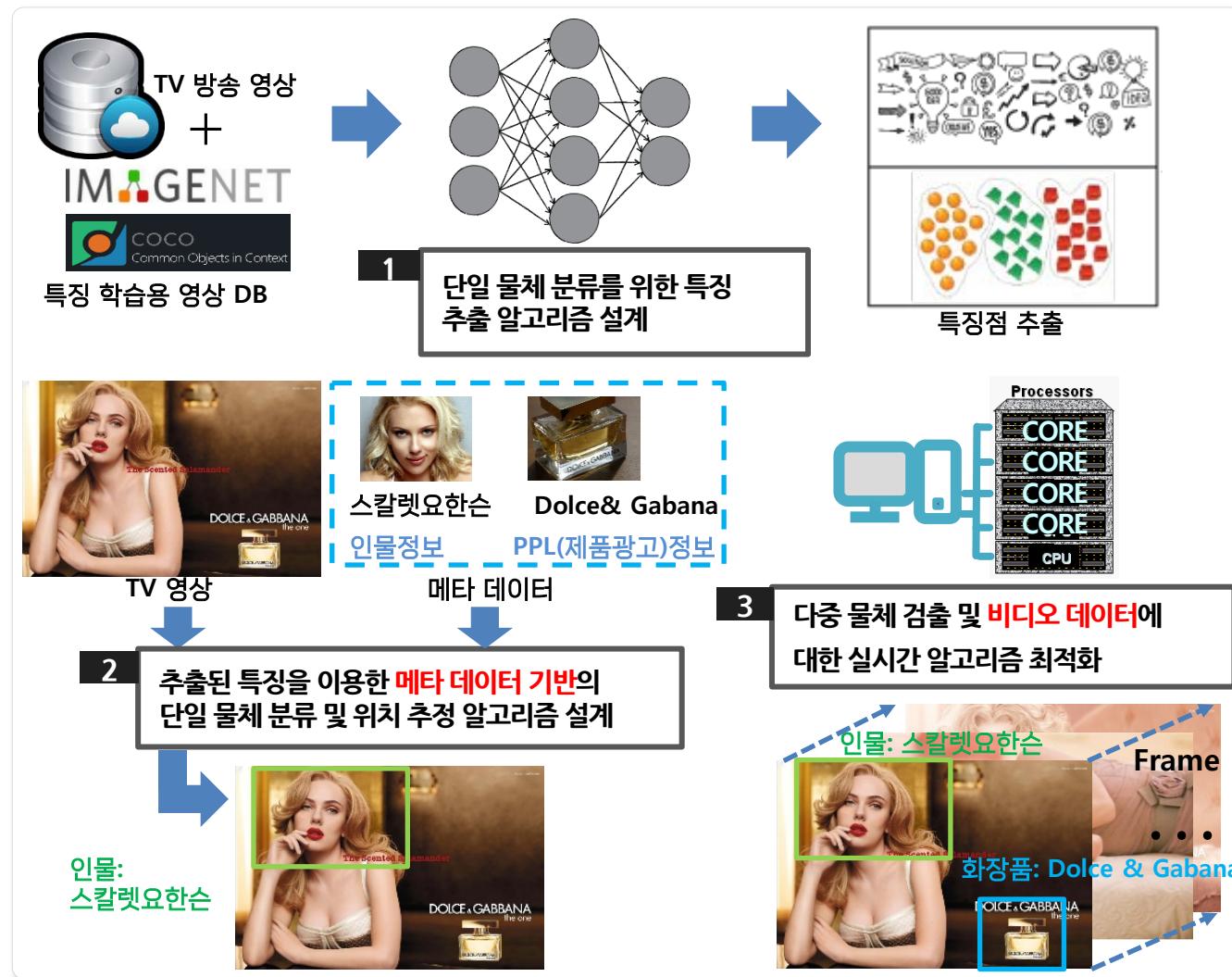
- 딥러닝 기반의 단일 물체에 대한 특징 추출 네트워크 개발
- TV 영상 콘텐츠에 특화된 데이터베이스 구축

2. 특징 기반의 메타데이터 정보를 활용한 단일 물체 분류 및 위치 추정 기술 연구

- 딥러닝 기반의 단일 물체 분류
- 분류된 물체의 정확한 위치 예측

3. 실시간 다중 물체 검출 및 위치 예측 기술 연구

- 다중 물체의 위치와 클래스를 종합하여 검출하는 기술 개발
- 비디오 데이터에 대한 안정적 물체 검출
- 실시간 동작을 위한 알고리즘 최적화



UHD 방송 환경에서 콘텐츠에 대한 시청자의 반응 및 의도 기반 미디어 인터랙션 기술

과학기술정보통신부, 2018-2020.

방송 콘텐츠 분석 및 인식

- 1차년도 연구 결과

1. TV 영상 컨텐츠에 특화된 데이터 베이스 구축

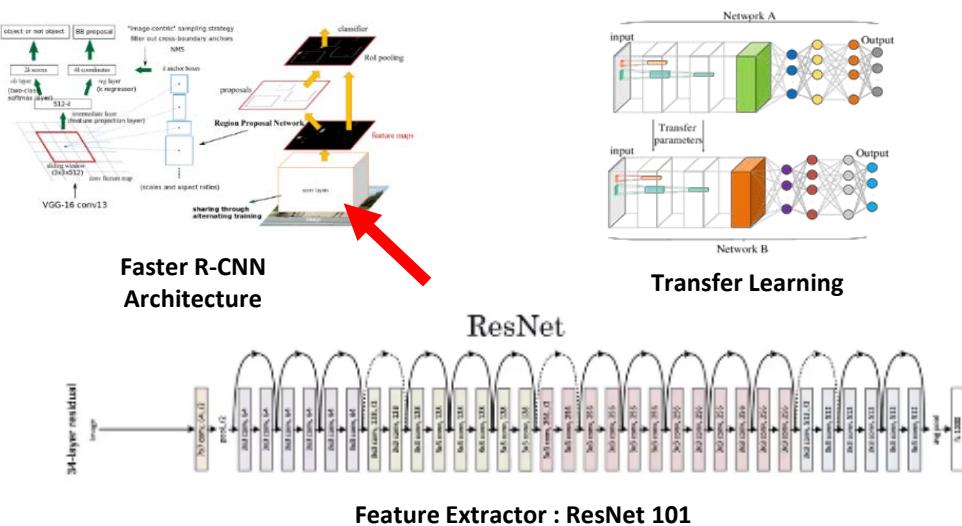
- 드라마 '미생' 영상 이미지를 취합, 데이터 구축
- '사람(Person)', '음료수병(Bottle)' 레이블
- 2, 3년차의 연구 활용을 위하여 주연 배우 2명과 주요 음료 간접 광고 제품에 대한 세부적인 Annotation 완료
- 자체적으로 만든 S/W tool로 Annotation 수행
- Annotation 파일은 xml 형식으로 저장



Dataset	Training Set	Test Set	Total
Images	580	74	654
Person Class	996	163	1159
Bottle Class	753	34	787

2. 딥러닝 기반의 단일 물체에 대한 특징점 추출 네트워크 개발

- 딥러닝 기반의 객체 검출 알고리즘인 Faster R-CNN(*) 구현
- Feature Extractor : ResNet 101(**) Layers
- MSCOCO Pretrained Model(***)로 Transfer Learning 기법 활용
- 개발 프로그래밍 언어 / API : Python / Google Tensorflow



* S.Ren, K.He, R.Grishick, J.Sun, "Towards real-time object detection with region proposed networks," ICCV 2015.

** K He, X Zhang, S Ren, J Sun, "Deep Residual Learning for Image Recognition", CVPR 2016

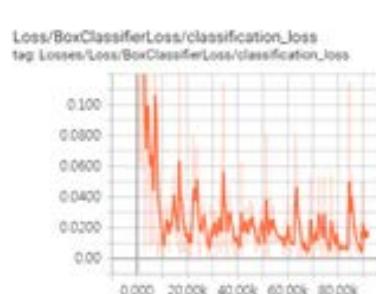
*** T Lin, M Maire, S Belongie, L Bourdev, R Girshick, "Microsoft COCO: Common Objects in Context", <https://arxiv.org/pdf/1405.0312.pdf>

UHD 방송 환경에서 콘텐츠에 대한 시청자의 반응 및 의도 기반 미디어 인터랙션 기술 과학기술정보통신부, 2018-2020.

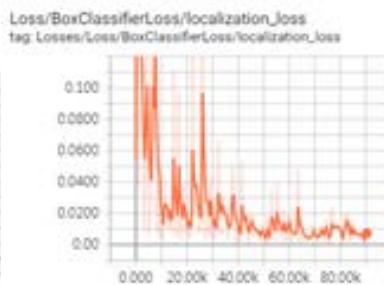
방송 콘텐츠 분석 및 인식 ▪ 1차년도 연구 결과

3. 연구 결과

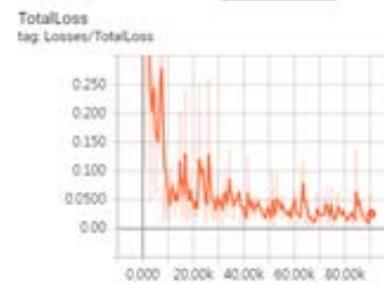
- 정의된 Loss 값이 수렴하며 정상적으로 학습되는 것을 관찰
- 기존 Faster R-CNN ResNet 101 model의 mAP 76.4를 정량적 목표치로 설정
- 본 연구 개발 모델이 방송 데이터에 대해서 mAP 81.8를 달성, 기존 논문 연구 모델보다 향상된 수치 달성



1) Box Classifier Loss



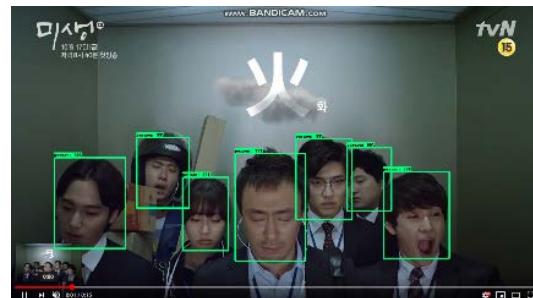
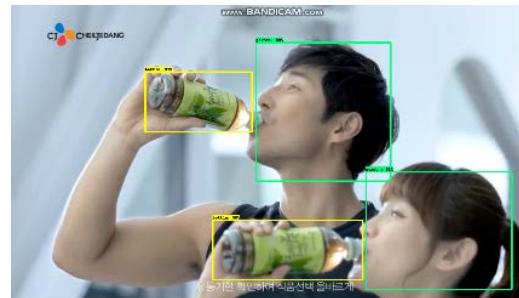
2) Box Localization Loss



3) Total Loss

Methods	Data Set	mAP	AP of Bottle	AP of Person
기존 모델	PASCAL, COCO	76.4	65.7	82.3
연구 개발 모델	방송 Data set	81.8	71.2	92.3

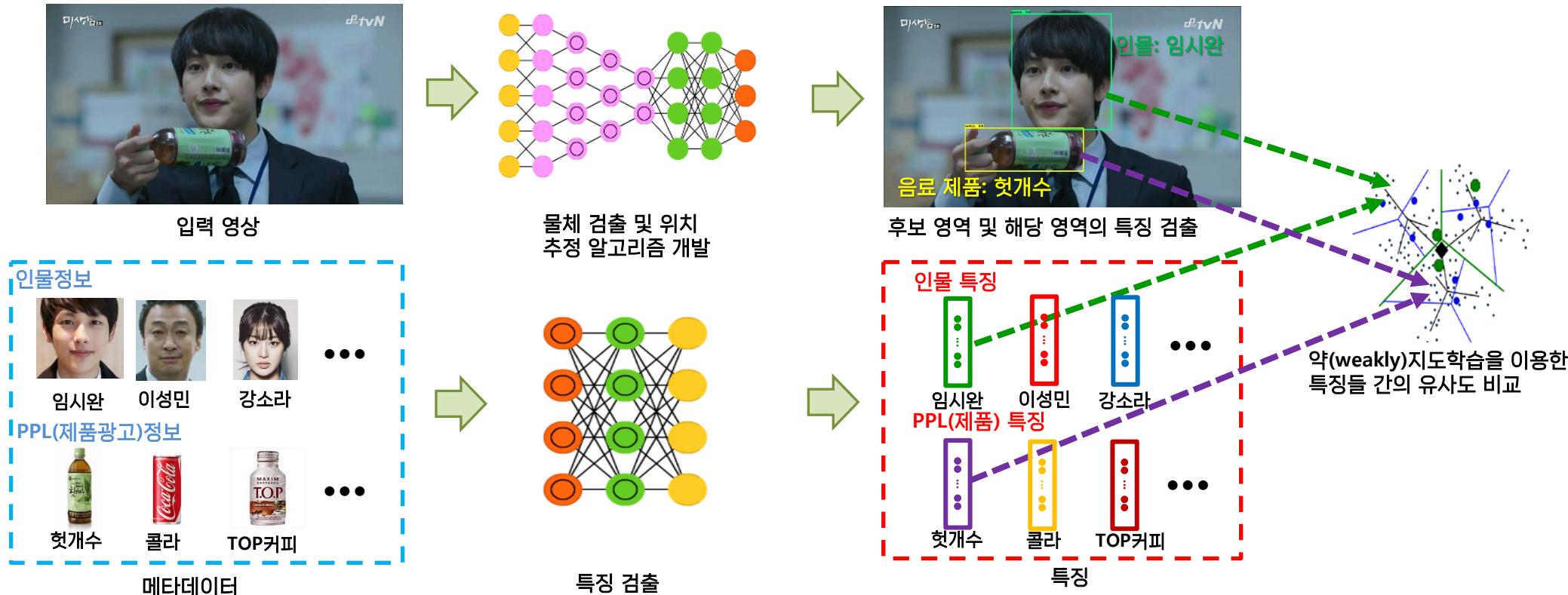
4. 테스트 결과 이미지



UHD 방송 환경에서 콘텐츠에 대한 시청자의 반응 및 의도 기반 미디어 인터랙션 기술 과학기술정보통신부, 2018-2020.

방송 콘텐츠 분석 및 인식

▪ 2차년도 연구 결과



1. 메타 데이터를 활용한 물체 분류 기술 개발

- 물체 분류 및 특징 기반의 메타데이터 정합 기술 개발
- 기존의 일반적인 물체에 대한 정의가 아닌 (영상 내 인물 및 대상들에 대한) 적은 양의 DB가 존재하는 관심 물체에 대하여 메타데이터에 기반한 약 지도 학습(weakly supervised learning)을 이용한 학습/검출

2. 분류된 물체의 정확한 위치 추정 기술 개발

- 분류된 물체의 위치에 대한 Bounding Box를 추정하는 기술

2차년도 정량적 목표	기준 수준	2차년도 목표
물체 검출 정확도(%)	76.4	76.5

※ 1차년도 목표치가 일반적인 범주 인식에 대한 것이었다면,
2차년도는 특정 Instance에 대한 인식 목표치

UHD 방송 환경에서 콘텐츠에 대한 시청자의 반응 및 의도 기반 미디어 인터랙션 기술 과학기술정보통신부, 2018-2020.

방송 콘텐츠 분석 및 인식

▪ 2차년도 연구 결과

1. 국내 방송인 Face DB 자체 구축



Years	Total Face DB	Face Classes	비고
1차년도	1,159 images	3 classes	<ul style="list-style-type: none">- 주로 드라마 '미생'에서 취득- 국내 연예인 3명 (classes): 임시완, 강소라, 이성민
2차년도	2,500 images	32 classes	<ul style="list-style-type: none">- 국내 연예인 총 32명(classes) 데이터 수집- 각 class 당 약 80여 장의 데이터 수집 완료- Website crawling을 통해 다양한 데이터 수집

2. PPL 상품(음료) 데이터베이스 구축



비타500, 컨디션, 콜라 등
다양한 Bottle 클래스 DB 확보

Years	Total Object DB	Bottle Classes	비고
1차년도	787 images	2 classes	<ul style="list-style-type: none">- 주로 드라마 '미생'에서 취득- Bottle 클래스: 헛개수, 조지아커피
2차년도	3,870 images	7 classes	<ul style="list-style-type: none">- Bottle 클래스 7개로 다양화- Website crawling을 통해 다양한 데이터 수집- 직접 동영상 촬영으로 데이터 수집

UHD 방송 환경에서 콘텐츠에 대한 시청자의 반응 및 의도 기반 미디어 인터랙션 기술 과학기술정보통신부, 2018-2020.

방송 콘텐츠 분석 및 인식

▪ 2차년도 연구 결과

▪ Test Set : 3,363 images

- 실제 방송 광고 영상을 기반으로 테스트셋 구축
- 배우 : 수지, 유재석, 정준하, 노홍철, 차승원
- 음료 : 비타500, 씨그램, 콜라



1. 인물 테스트 결과



	2차년도 목표	2차년도 달성	유재석	노홍철	정준하	차승원	수지
mAP(%)	76.5	77.22	82.85	74.49	76.37	77.68	77.22

2. PPL 상품(음료) 테스트 결과

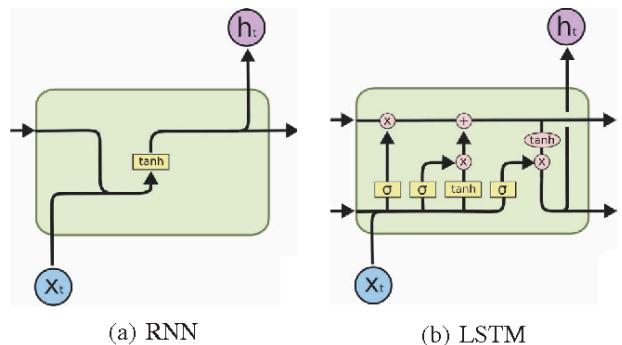
	2차년도 목표	2차년도 달성	코카콜라	씨그램	비타500
mAP(%)	76.5	78.88	82.85	74.49	76.37

UHD 방송 환경에서 콘텐츠에 대한 시청자의 반응 및 의도 기반 미디어 인터랙션 기술 과학기술정보통신부, 2018-2020.

방송 콘텐츠 분석 및 인식

- 3차년도 연구 결과

3차년도



RNN(Recurrent Neural Networks)과 LSTM(Long Short-Term Memory Models) 알고리즘 개발 및 활용



<3차년도 연구목표 및 내용>

- 영상 기반의 다중 물체 검출 및 위치 예측 기술 개발
- 비디오를 위한 실시간 검출 알고리즘 개발 및 시스템 최적화

<기술적 차별성>

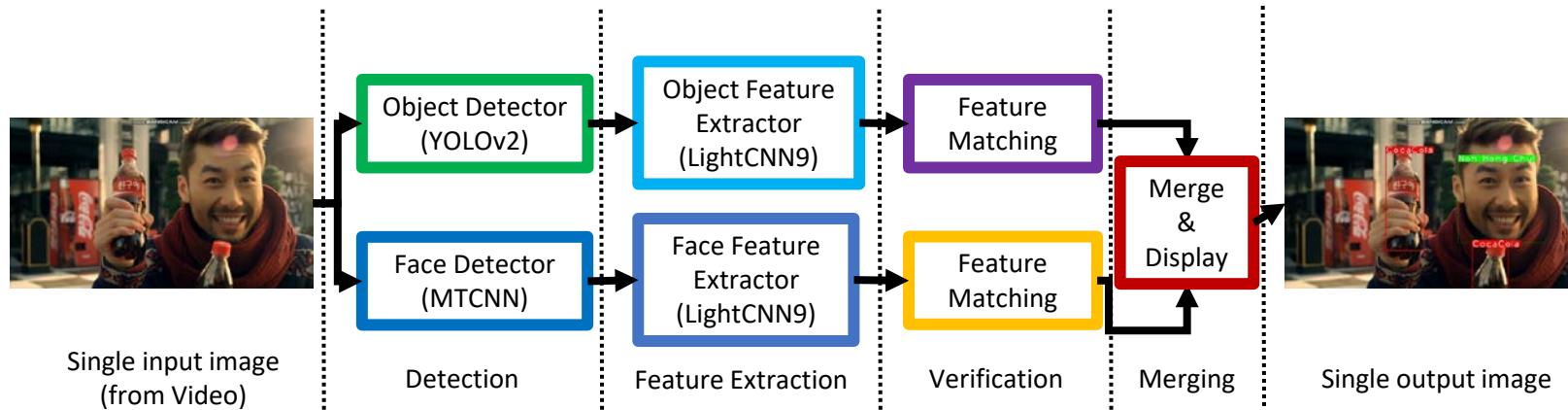
- 비디오 데이터에 특화된 실시간 물체 분류 및 위치 추정

정량적 목표	2차년도 목표 수준	3차년도 목표 수준
물체 검출 정확도	76.50%	77%

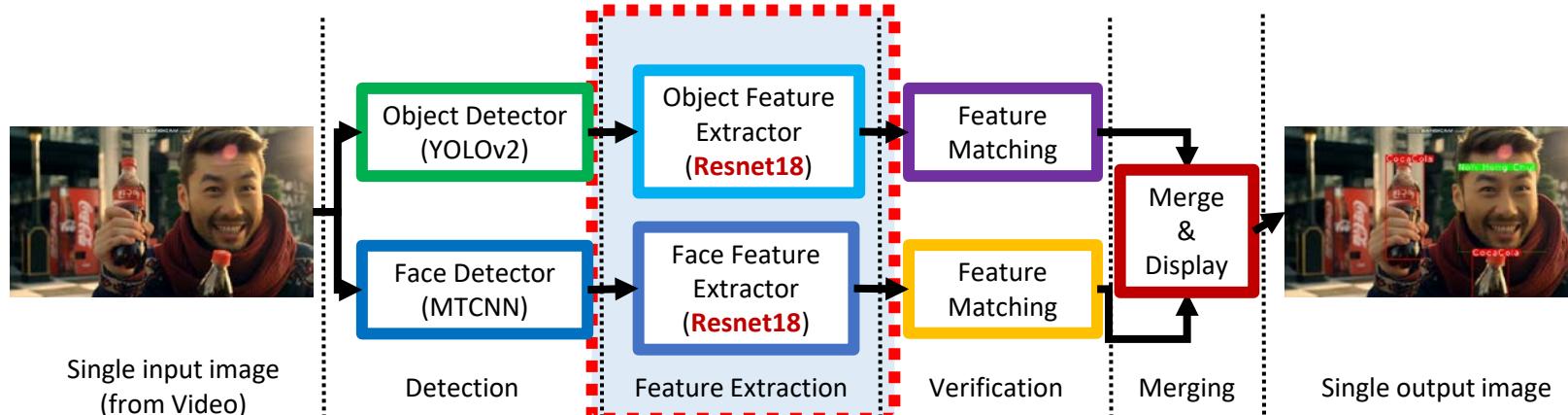
UHD 방송 환경에서 콘텐츠에 대한 시청자의 반응 및 의도 기반 미디어 인터랙션 기술 과학기술정보통신부, 2018-2020.

방송 콘텐츠 분석 및 인식

- 3차년도 연구 결과



2nd Year: Training & Test with LightCNN 9 model



3rd Year: Training & Test with Resnet 18 + Arcface module

UHD 방송 환경에서 콘텐츠에 대한 시청자의 반응 및 의도 기반 미디어 인터랙션 기술 과학기술정보통신부, 2018-2020.

방송 콘텐츠 분석 및 인식

▪ 3차년도 연구 결과

- LightCNN 9 (2차년도 모델) vs Resnet 18 (3차년도 모델) PPL 상품(음료) mAP 성능 비교
 - 동일한 Training DB(자체 음료 Dataset)로 학습
 - 2차년도 광고 영상 테스트 셋에 대한 성능 평가 결과

1. 인물 테스트 결과

Years	Model	mAP(↑)	mAP(↑)						
			강소라	강하늘	김대명	박해준	변요한	이성민	임시완
2차년도	LightCNN 9	73.96%	57.86%	89.68%	88.06%	93.37	42.56%	74.52%	71.69%
3차년도	Resnet 18	98.08% (+24.12%)	91.85%	91.99%	98.43%	99.48%	82.23%	100.00%	92.47%

2. PPL 상품(음료) 테스트 결과

Years	Model	mAP(↑)	AP(↑)		
			CocaCola	Seagram	Vita500
2차년도	LightCNN 9	78.88%	76.15%	80.23%	80.26%
3차년도	Resnet 18	90.28%(+14.22%)	91.88%	87.98%	90.98%

UHD 방송 환경에서 콘텐츠에 대한 시청자의 반응 및 의도 기반 미디어 인터랙션 기술 과학기술정보통신부, 2018-2020.

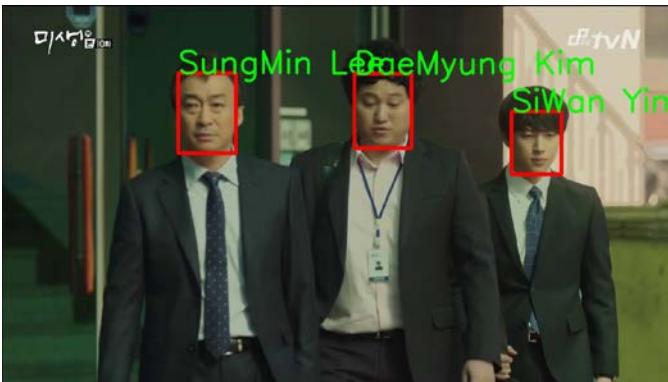
방송 콘텐츠 분석 및 인식

- 3차년도 연구 결과

Test Video



Result



Thank You !



Any Questions?