
Special Lecture

CHRIST
(DEEMED TO BE UNIVERSITY)
BANGALORE, INDIA

Event: Special Lecture

Date: 16th November 2022

Venue: Online

High-performance portrait segmentation using the ensemble of deep-learning models

Associate Professor, Yong-Woon Kim

Department of Computer Science and Engineering
CHRIST (Deemed to be University)

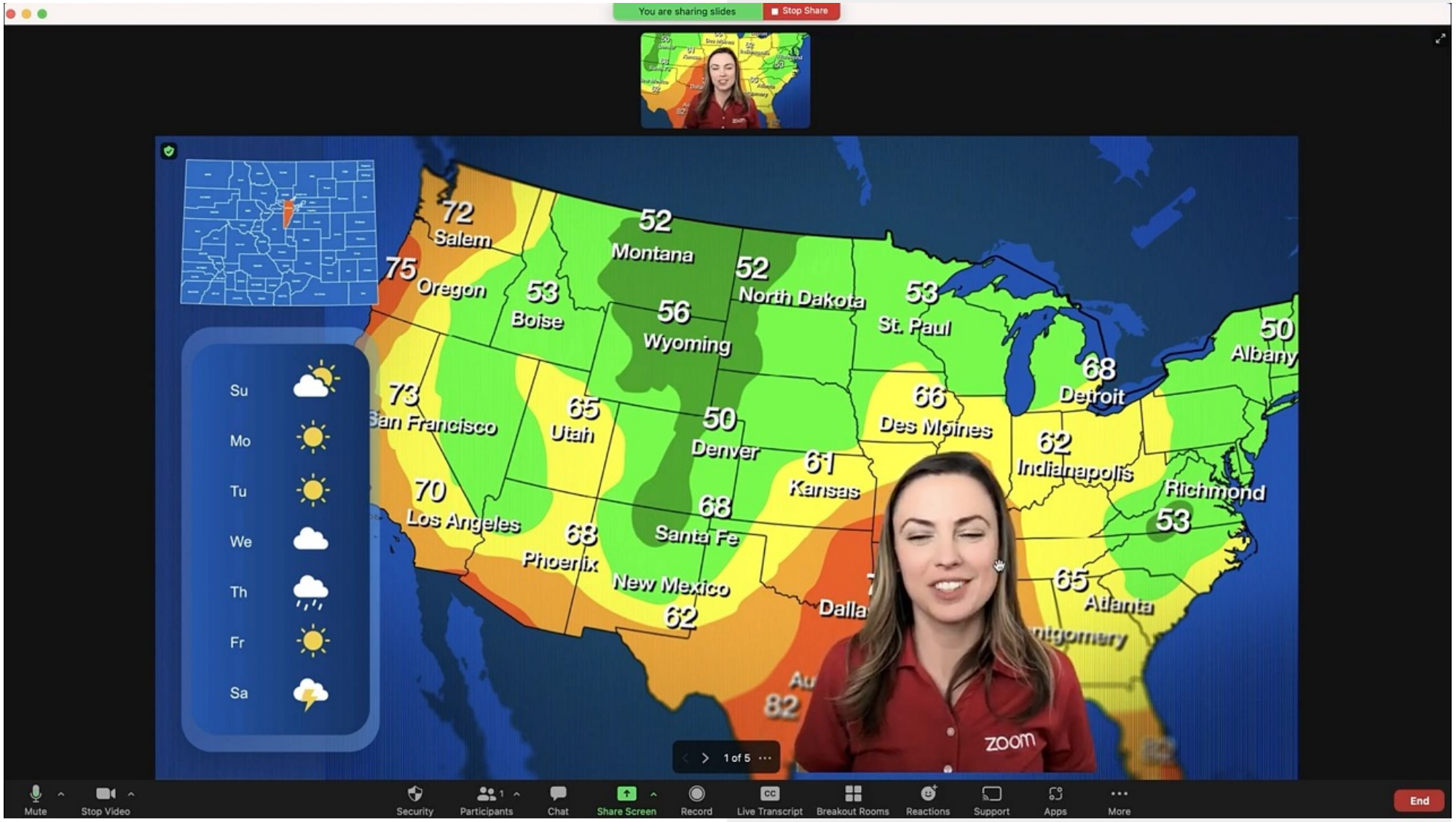
What is

PORTRAIT SEGMENTATION

Background change of camera app

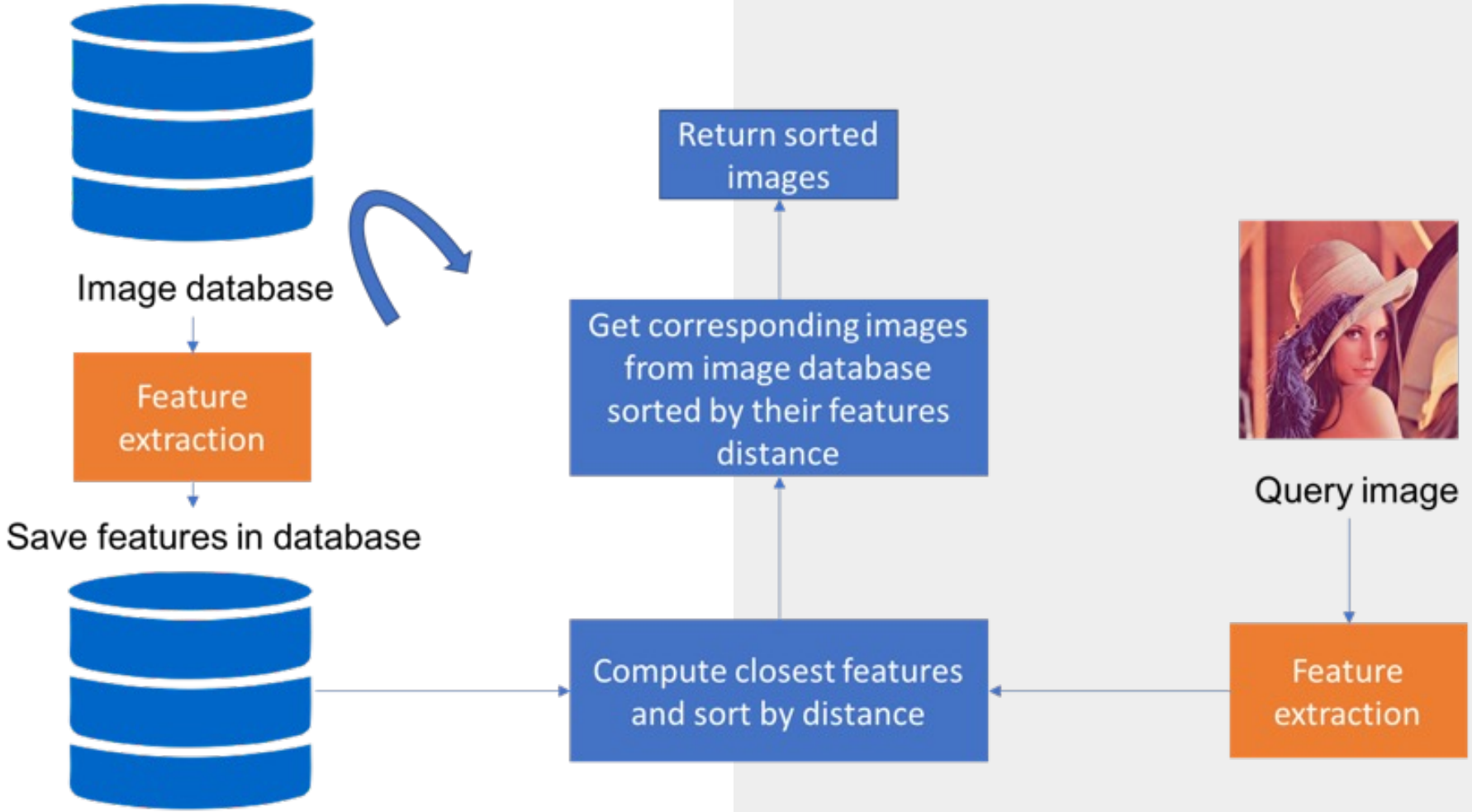
Virtual background function of Zoom meeting

Virtual background function of Zoom meeting



Robot-Human Interaction

Content based image retrieval



Review of related works

Related Works	Occurrence
A. More than 120 related works and more than 1,900 references were reviewed.	
B. Studies on Portrait Segmentation (PS) using a single Deep-Learning based semantic Segmentation Model (DSM).	2016 ~ (14 papers until 2021)
C. Studies on object segmentation using an ensemble of DSM.	2016 ~ (5 papers until 2021)
D. Studies on portrait video segmentation using a hybrid method of a DSM and image processing algorithms.	2019 (1 paper)
E. Studies on PS using the ensemble of DSMs.	Not reported

Motivation, requirements, challenges and gaps

Motivation	• Many applications: selfie camera apps, online conferencing solutions, content-based image retrieval, robot-human interaction, etc.
Requirements	• High-accuracy and/or high-speed portrait segmentation for various applications.
Challenges	• High-performance and real-time PS remains a challenging problem even with recent developments on image segmentation using single DSMs. • Several studies on image segmentation using the ensemble of DSMs show better accuracy compared to single DSMs. However, the ensemble approach is not fit well when we directly apply to portrait video segmentation.

Gaps	• A good number of studies show that:
	  ○ DSM is a reasonable choice for the PS (14 related works)
	  ○ Ensemble of multiple DSMs can improve the precision of segmentation (5 related works)
	• However, the PS using the ensemble of multiple DSMs have not been studied.

Objectives and approaches

	OBJECTIVE-1	OBJECTIVE -2
	PS for selfie photos using ensemble method	PS for portrait videos using ensemble method
HIGH-ACCURACY	*Ensemble of Heterogeneous DSMs*	*Ensemble of N-Frames*
HIGH-SPEED	*Optimal combination*	*Rotation method*
	Downscaling of images	*Downscaling of images*

Comparison of approaches

		Related Works	This Work
OBJ-1	Speed	• Optimization of DSM architecture (11 papers).	• Optimal combination of DSMs.
	Accuracy	• Optimization of DSM architecture. • Ensemble of DSMs for medical or scene segmentation (5 papers).	• Ensemble of DSMs for portrait image segmentation.
OBJ-2	Speed	• Optimization of DSM architecture (3 papers).	• Rotation of multiple DSMs.
	Accuracy	• Optimization of DSM architecture.	• N-Frames ensemble of DSMs.

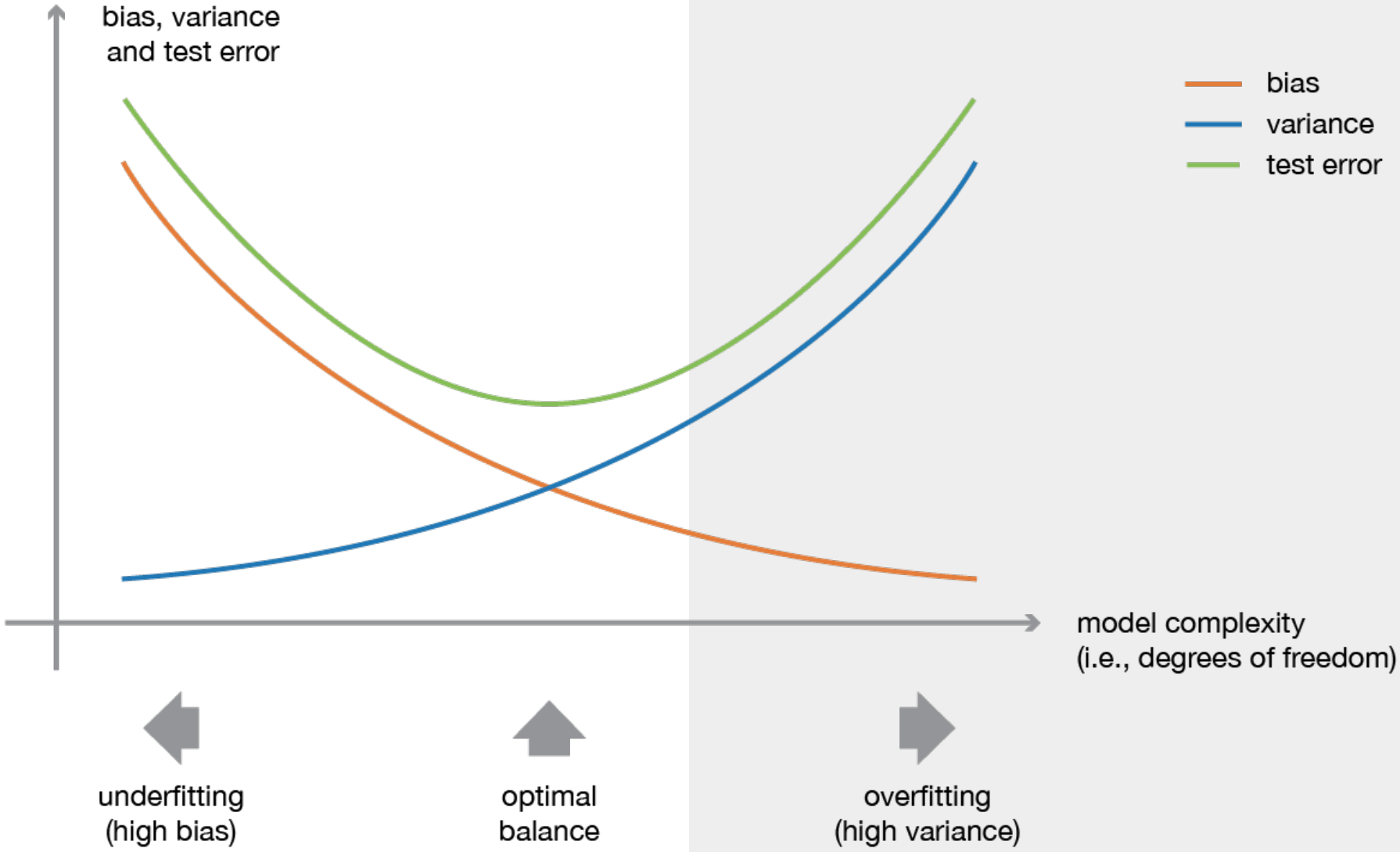
Measurement metrics

Measurement	Metrics
Accuracy	• Intersection over Union (IoU): $IoU = \frac{A \cap B}{A \cup B}$
Variance error	• IoU standard deviation
Bias error	• False Negative Rate: $FNR = \frac{FN}{FN+TP}$ • False Division Rate: $FDR = \frac{FP}{FP+TP}$
Efficiency	• Memory Efficiency Rate: $MER = \frac{M}{IoU}$ • Computing power Efficiency Rate: $CER = \frac{C}{IoU}$

Ensemble Model

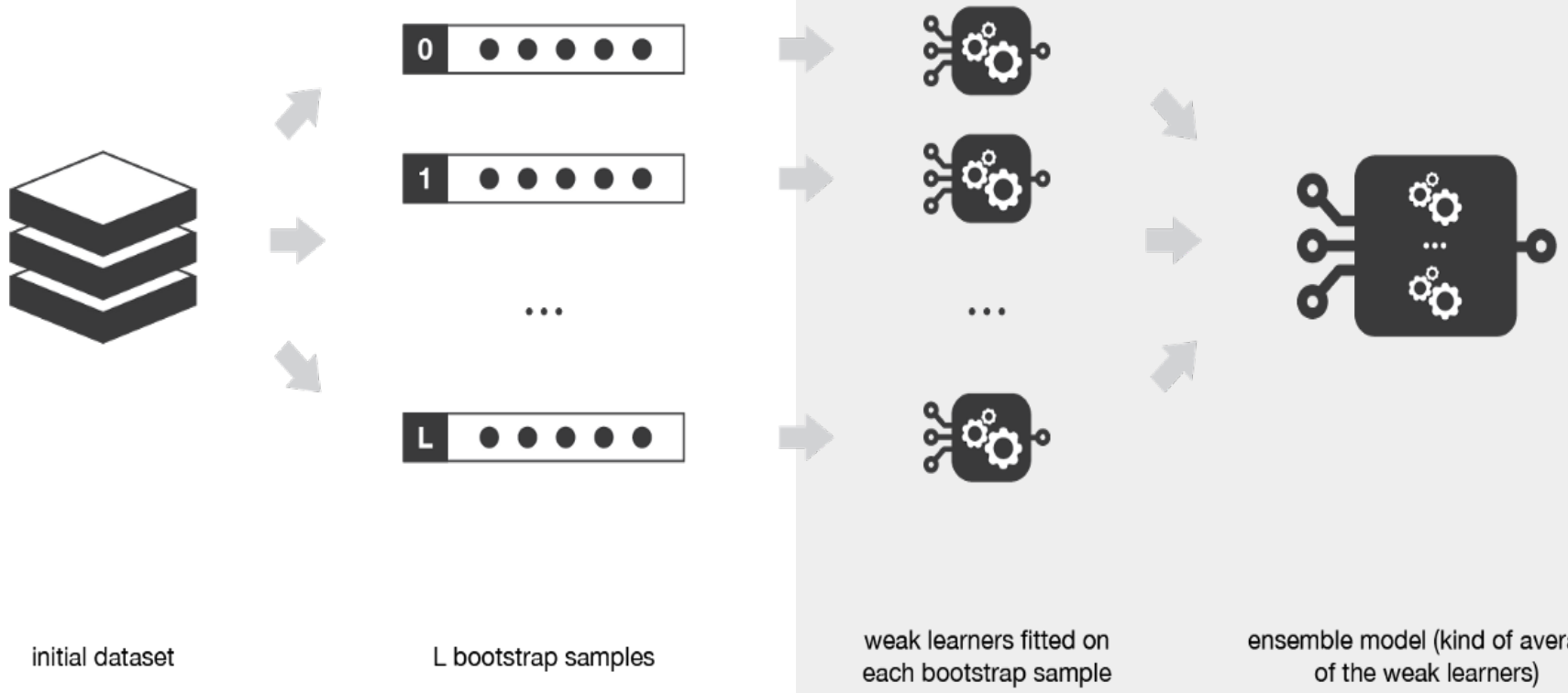
1. Bagging Ensemble
2. Boosting Ensemble
3. Stacking Ensemble

General Issues of Machine Learning: **Bias, Variance Errors and Overfitting**



(Source: https://towardsdatascience.com/ensemble-methods-bagging-boosting-and-stacking-c9214a10a205)

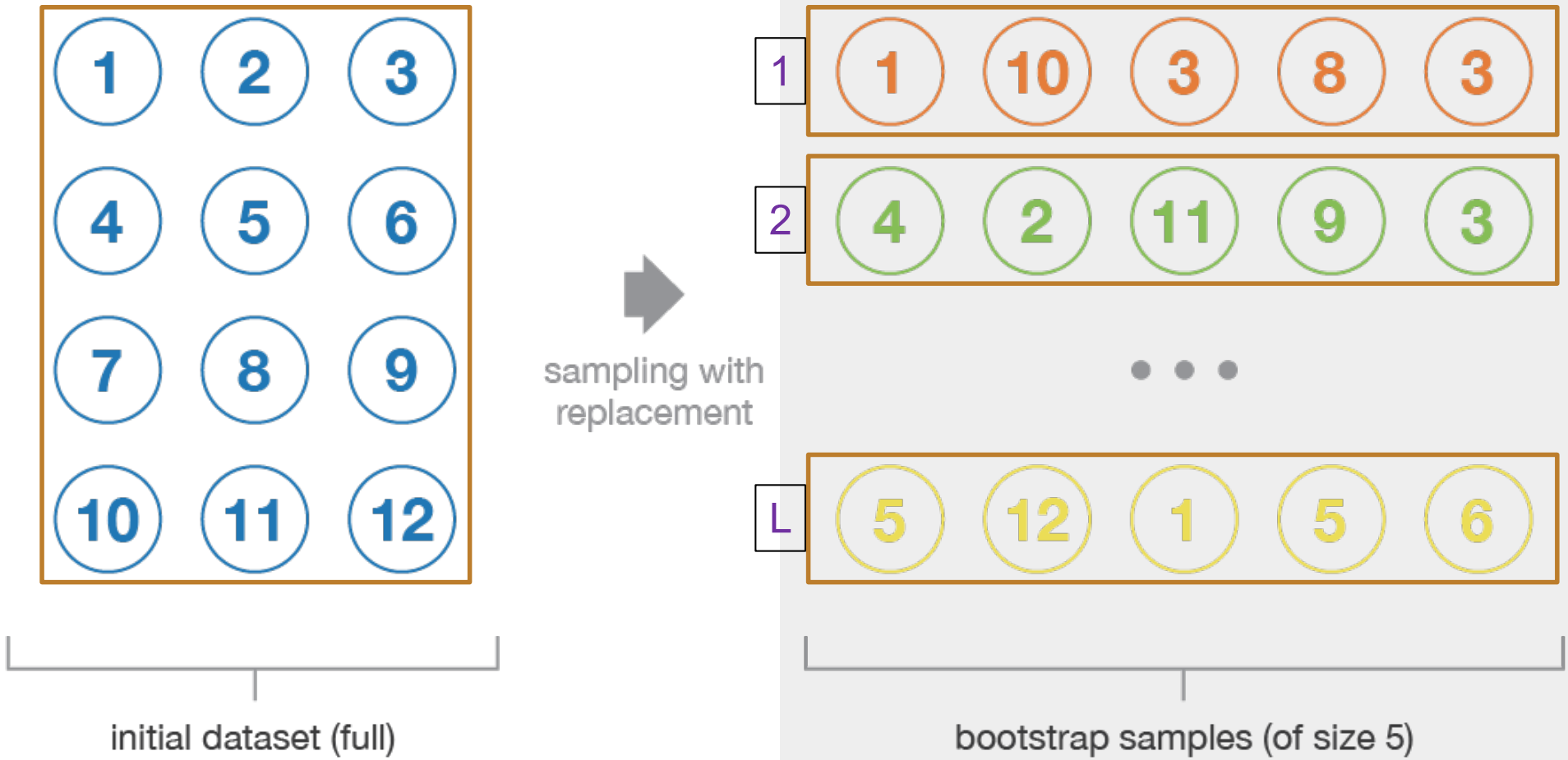
1) Bagging Ensemble



- It can reduce variance error
- It is suitable for a weak learner having low bias and high variance error

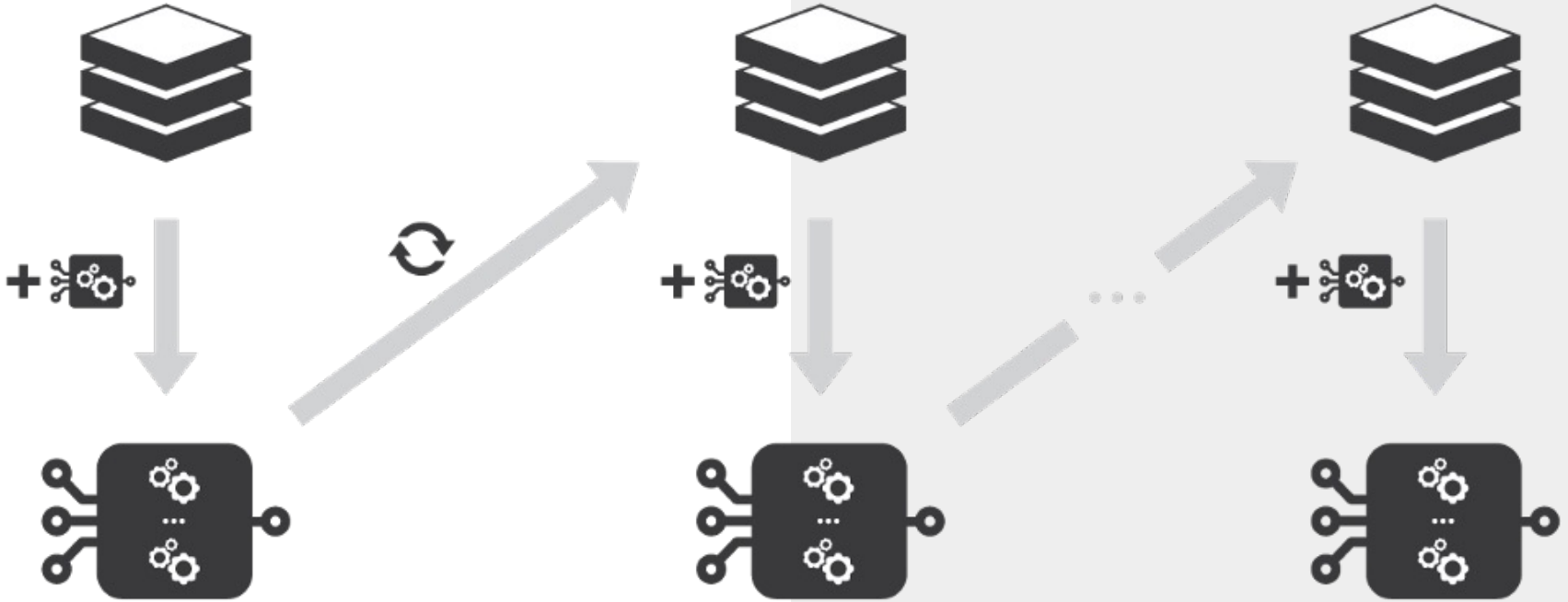
(Source: https://towardsdatascience.com/ensemble-methods-bagging-boosting-and-stacking-c9214a10a205)

Bootstrap samples



(Source: https://towardsdatascience.com/ensemble-methods-bagging-boosting-and-stacking-c9214a10a205)

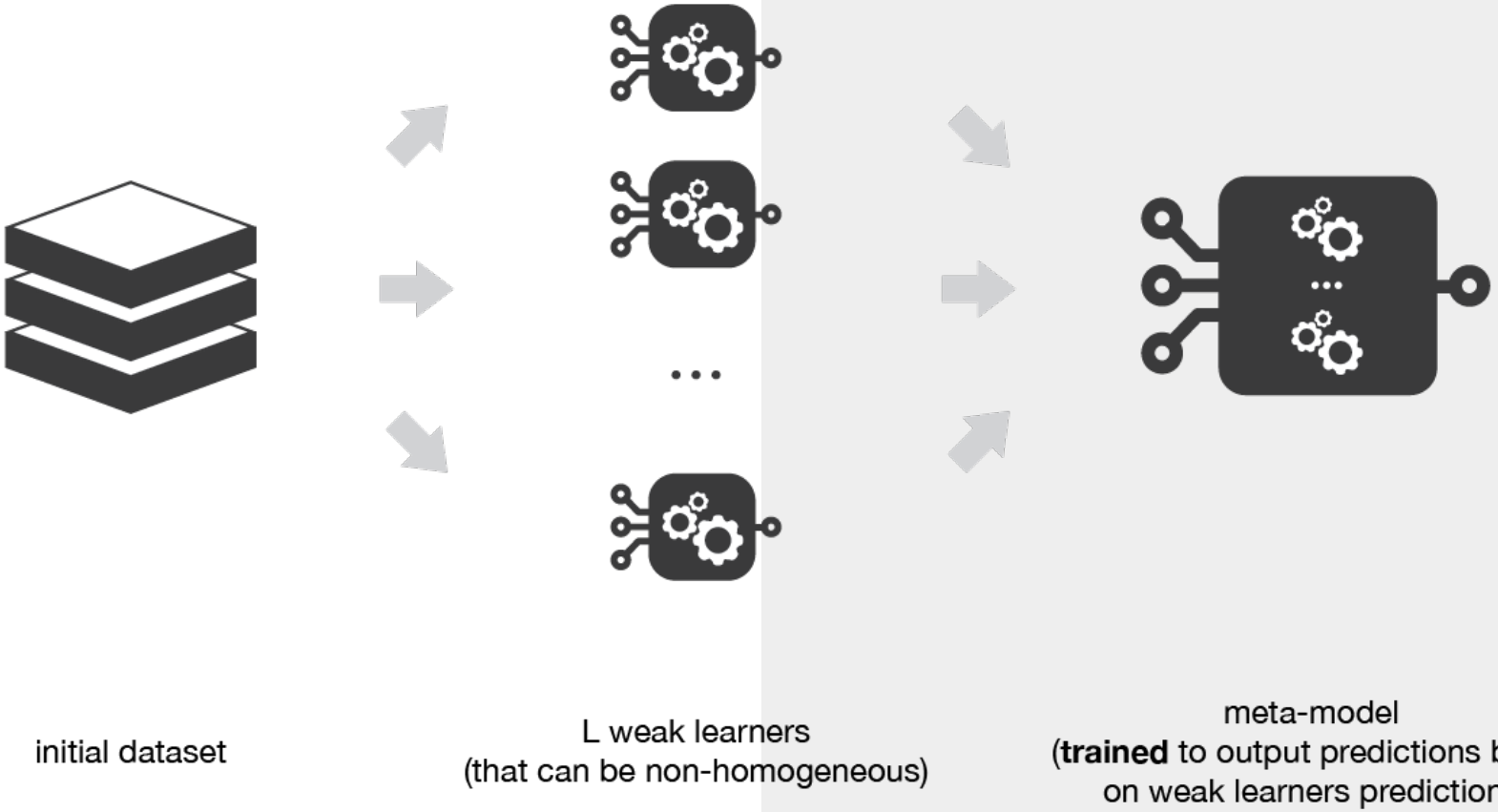
2) Boosting Ensemble



- It can reduce bias error
- It is suitable for a weak learner having low variance and high bias error

(Source: https://towardsdatascience.com/ensemble-methods-bagging-boosting-and-stacking-c9214a10a205)

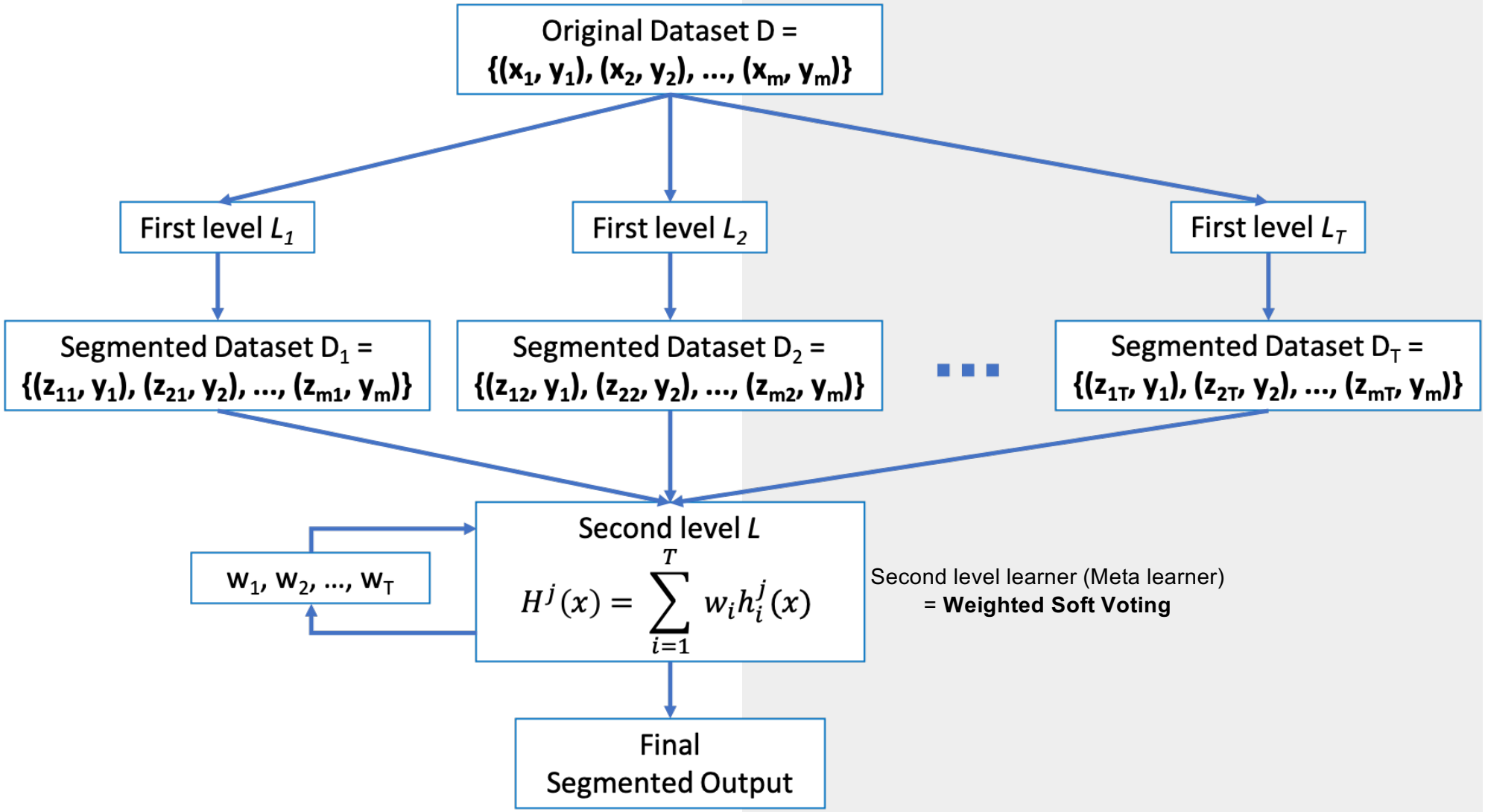
3) Stacking Ensemble



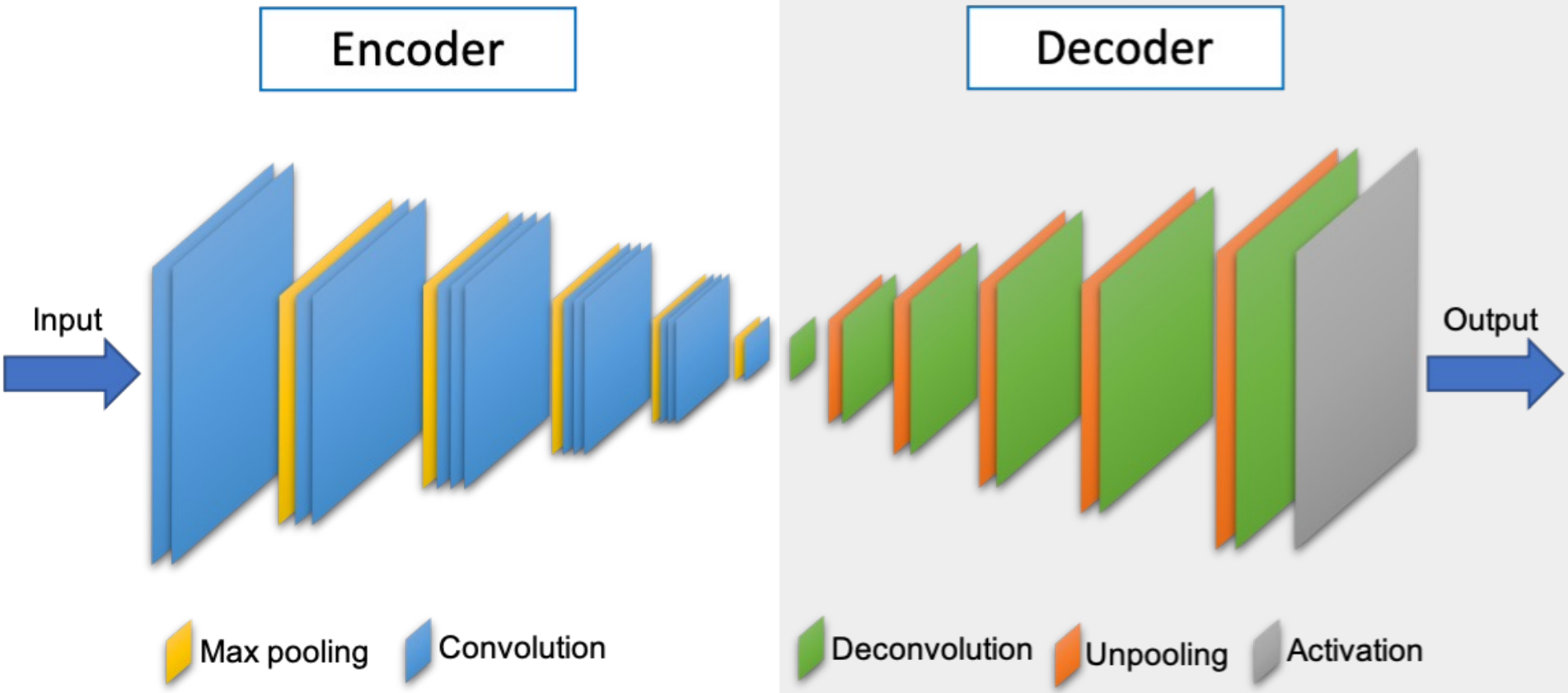
- It can improve accuracy
- It uses heterogeneous weak learners

(Source: https://towardsdatascience.com/ensemble-methods-bagging-boosting-and-stacking-c9214a10a205)

Stacking Ensemble for PS

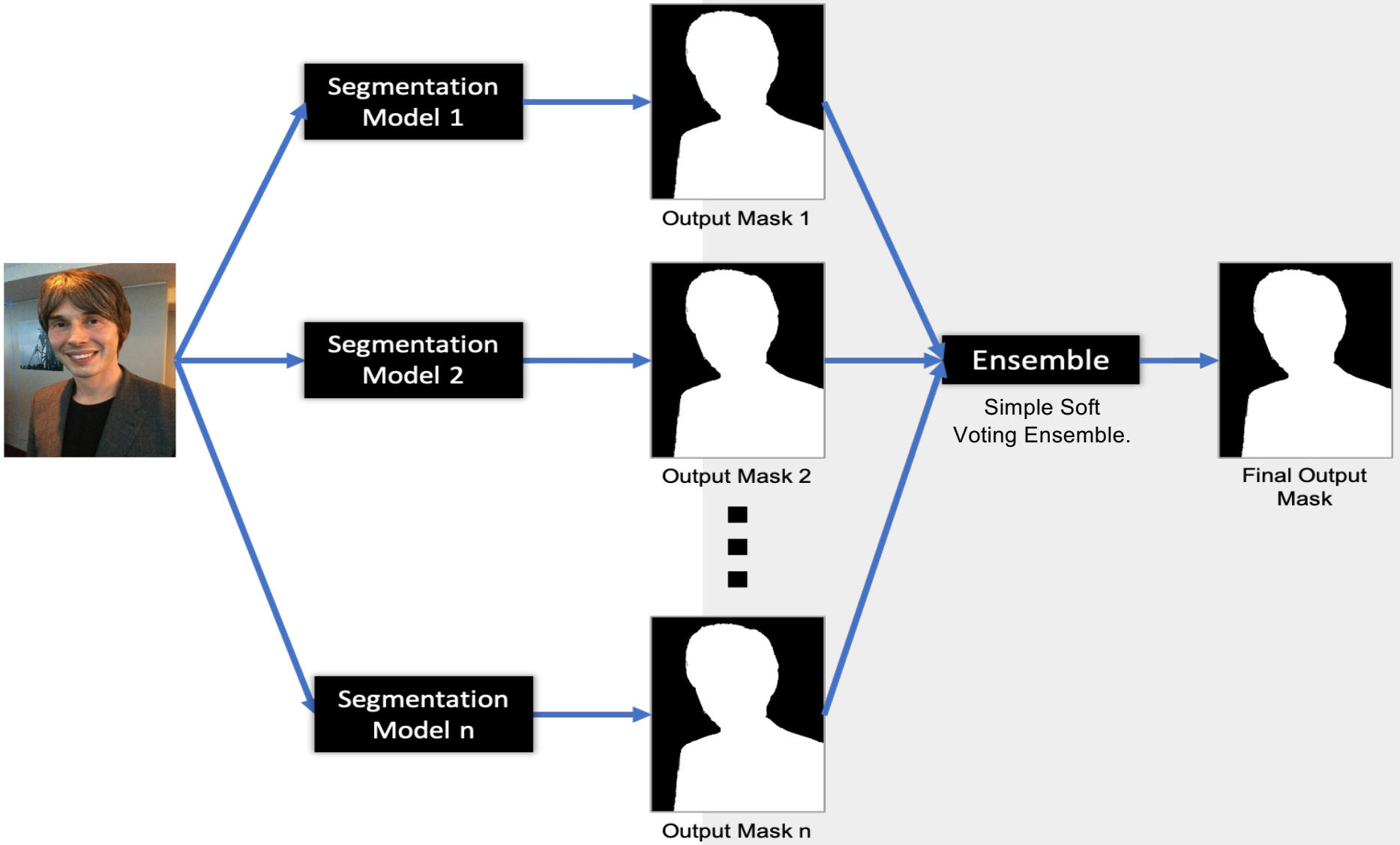


Deep-Learning based Segmentation Model (DSM)



[Approach-1]
PS using ensemble of Heterogeneous DSMs

PS using Stacking Ensemble

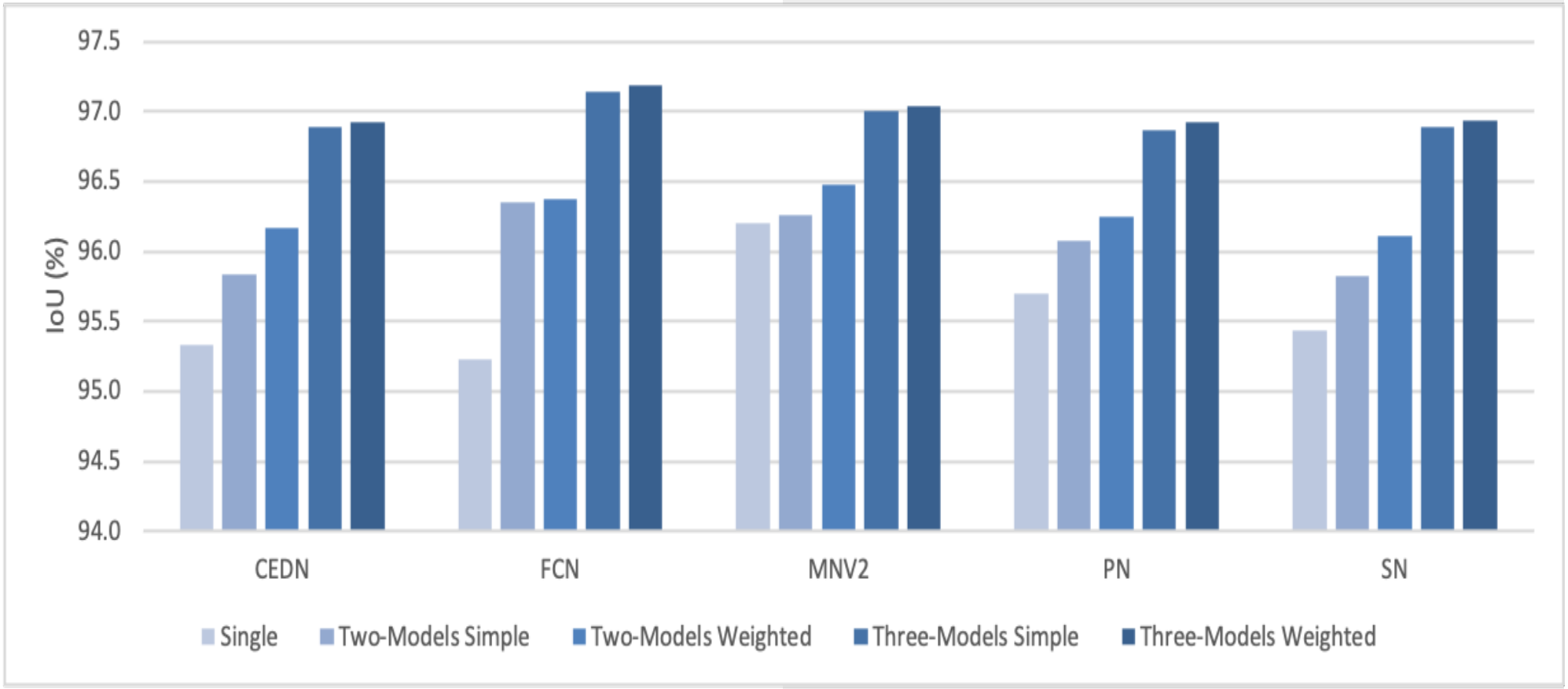


Segmentation Models used for Experiment

DSM	Training Dataset	Resolution	Testing Dataset
CEDN	AISegment	224x224	EG1800+CDI
FCN	PASCAL VOC 2011	500x500	EG1800+CDI
MNV2 (MobileNetV2)	Custom	128x128	EG1800+CDI
MNV3 (MobileNetV3)	Custom	224x224	EG1800+CDI
PN (PortraitNet)	EG1800	224x224	EG1800+CDI
SN (SINet)	EG1800	224x224	EG1800+CDI

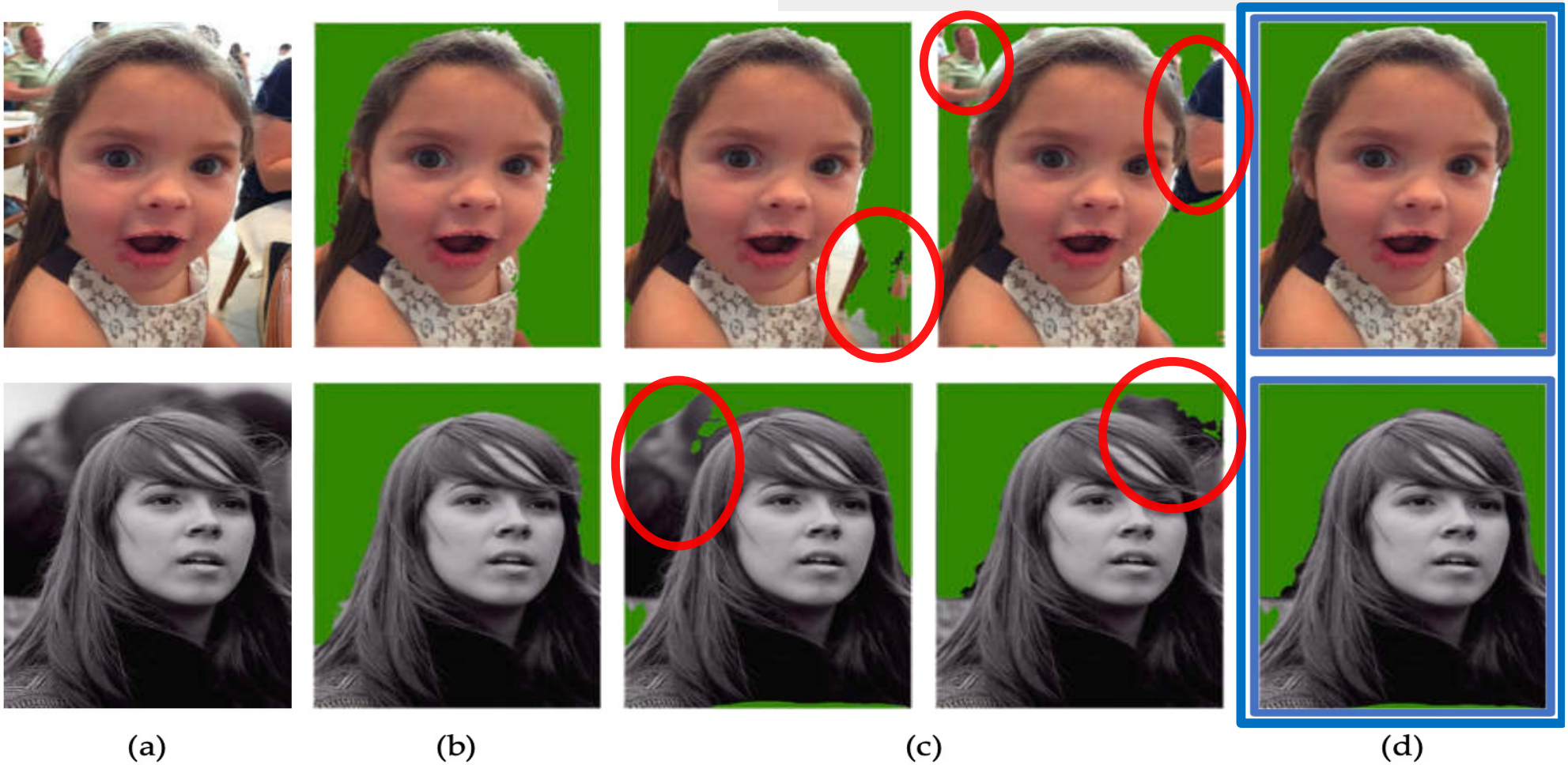
- DSM: Deep-Learning based semantic segmentation model.

IoU (%) Comparision



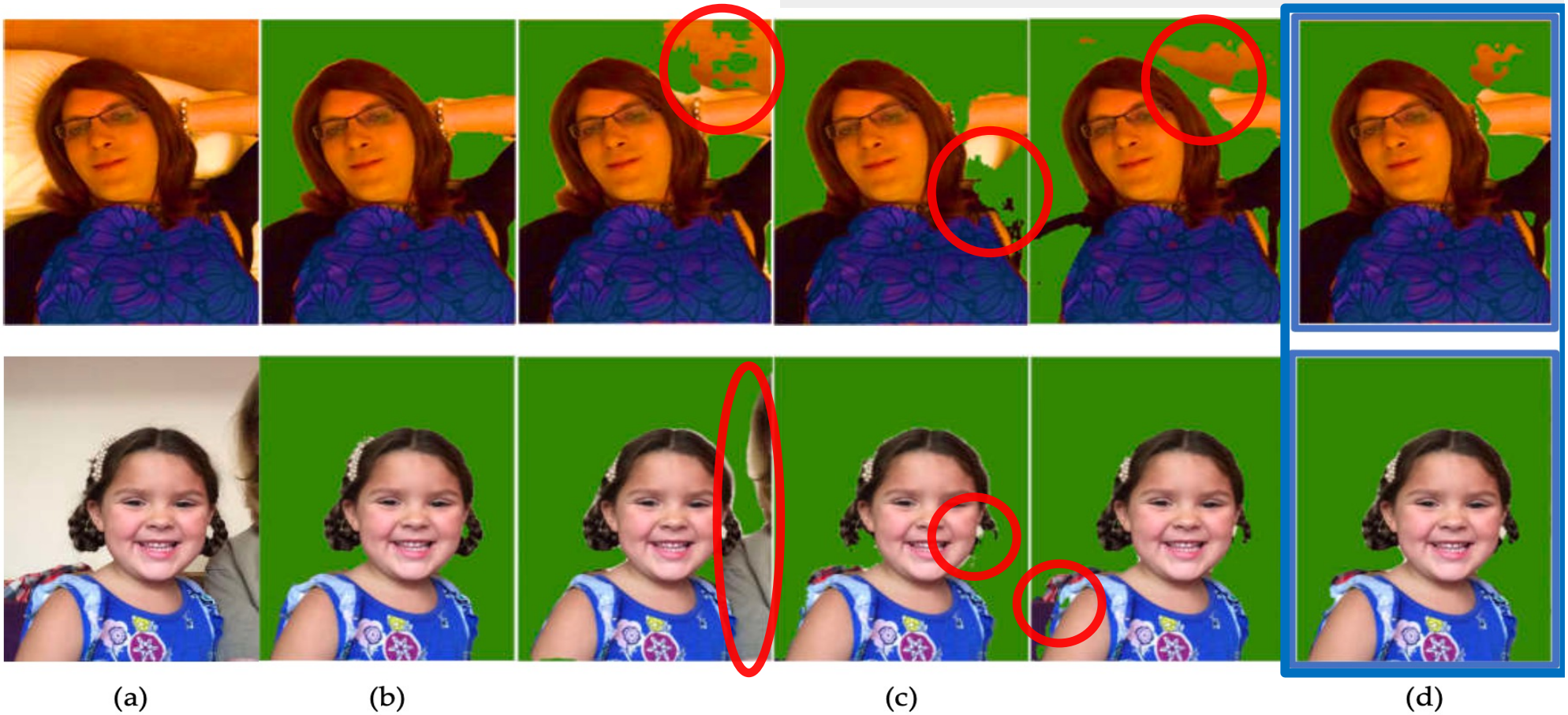
- IoU (Intersection over Union): A metric to measure the accuracy of segmentation.

Segmentation result of selfie photos



(a) Input image; (b) Expected output; (c) output of Segmentation models;
(d) output of proposed Two-Models Ensemble.

Segmentation result of selfie photos



(a) Input image; (b) Expected output; (c) output of Segmentation models;
(d) output of proposed Three-Models Ensemble.

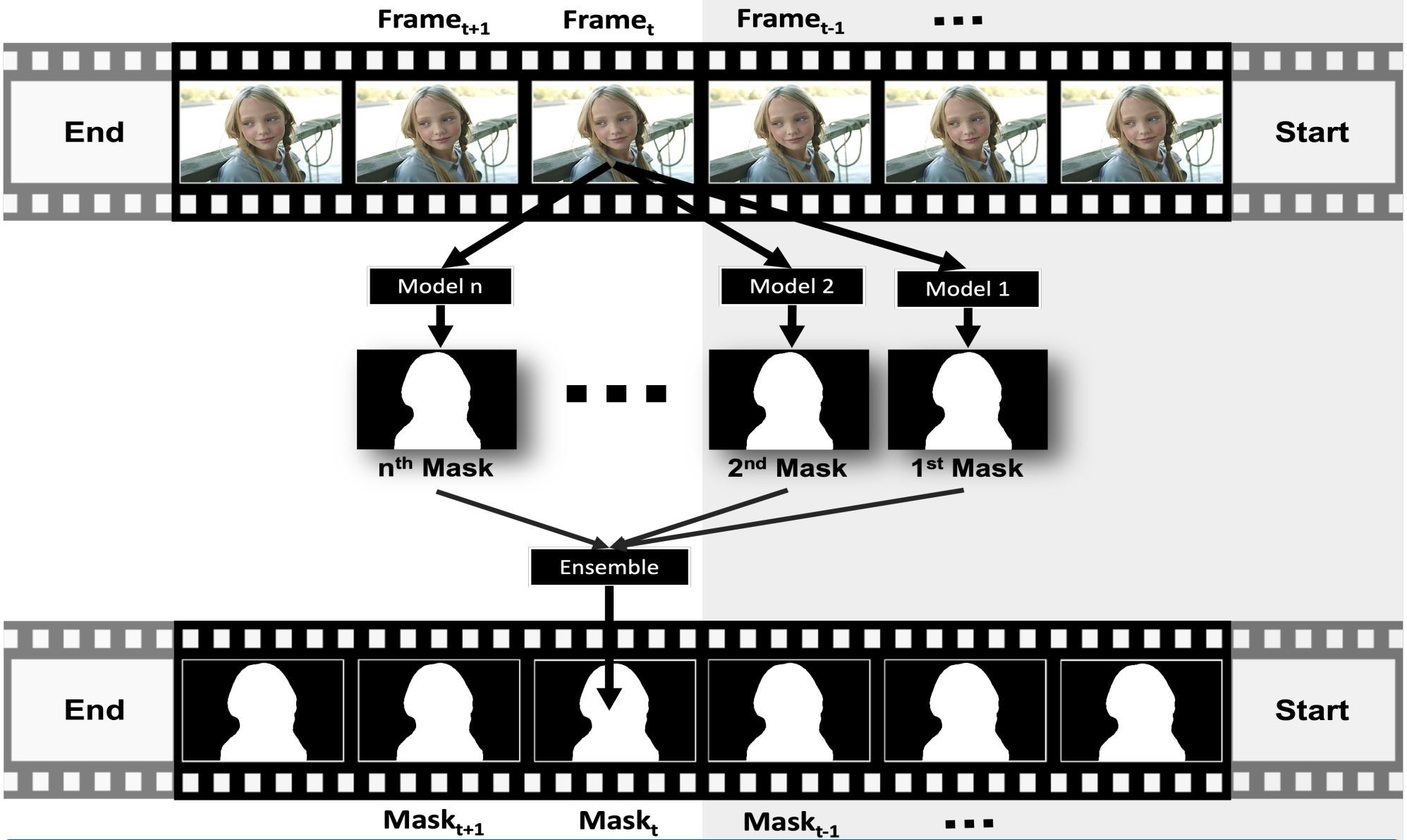
Efficiency Comparison

Model	Params (M)	FLOPs (G)	IoU (%)	MER	CER
SN	0.087	0.15	95.4326	0.091	0.157
PN	2.115	0.209	95.6992	2.210	0.218
MNV3	1.192	2.3724	94.7597	1.258	2.504
MNV2	3.625	7.227	96.2082	3.768	7.512
FCN	134.27	62.89	95.2268	141.000	66.042

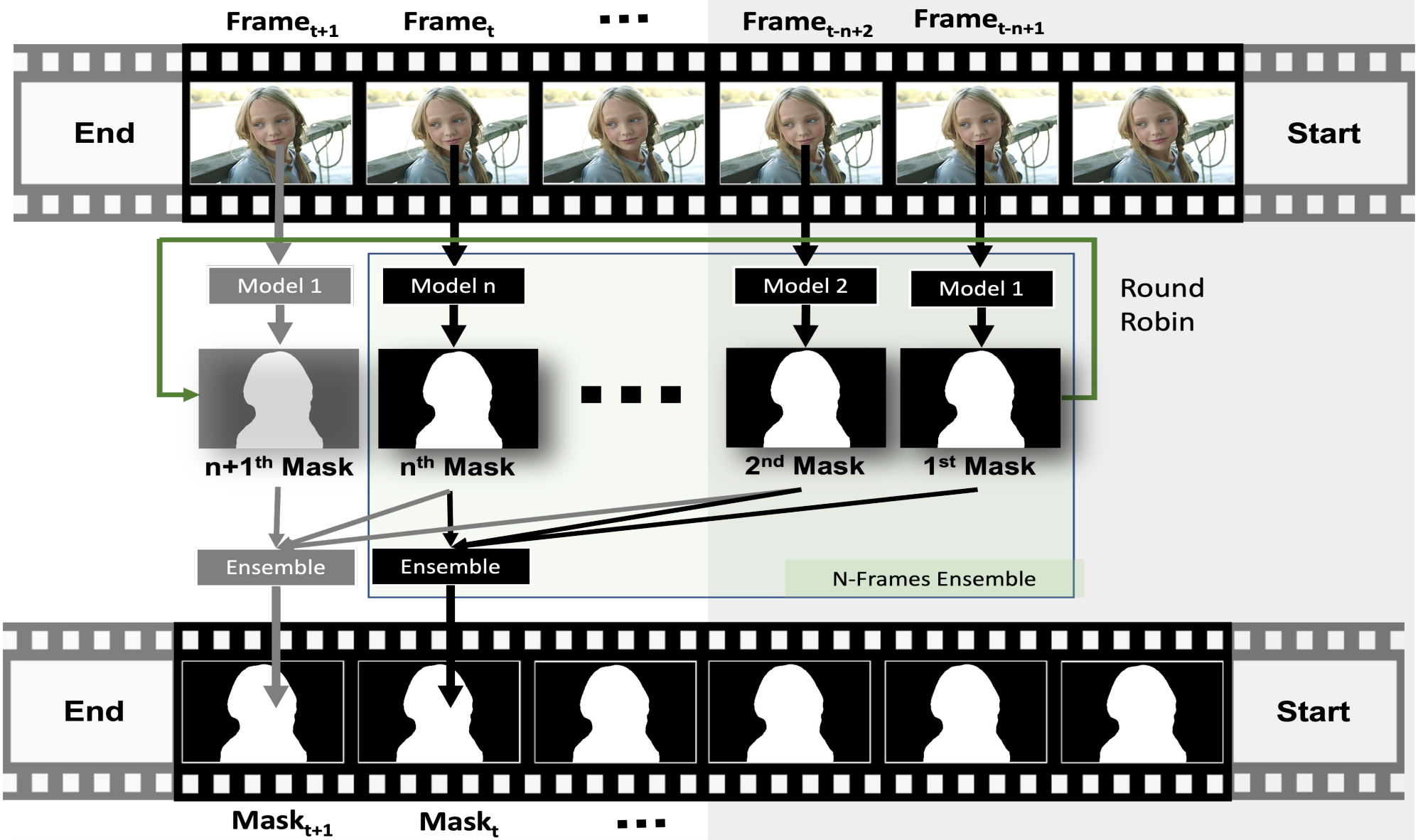
Model	Params (M)	FLOPs (G)	IoU (%)	MER	CER
MNV3 + PN + SN	3.394	2.7314	96.5053	3.517	2.830
MNV2 + PN + SN	5.827	7.586	96.7476	6.023	7.841
MNV3 + MNV2 + SN	4.904	9.7494	96.6016	5.077	10.092

[Approach-2]
PVS (Portrait Video Segmentation) using N-Frames ensemble of DSMs

General N-Models Ensemble



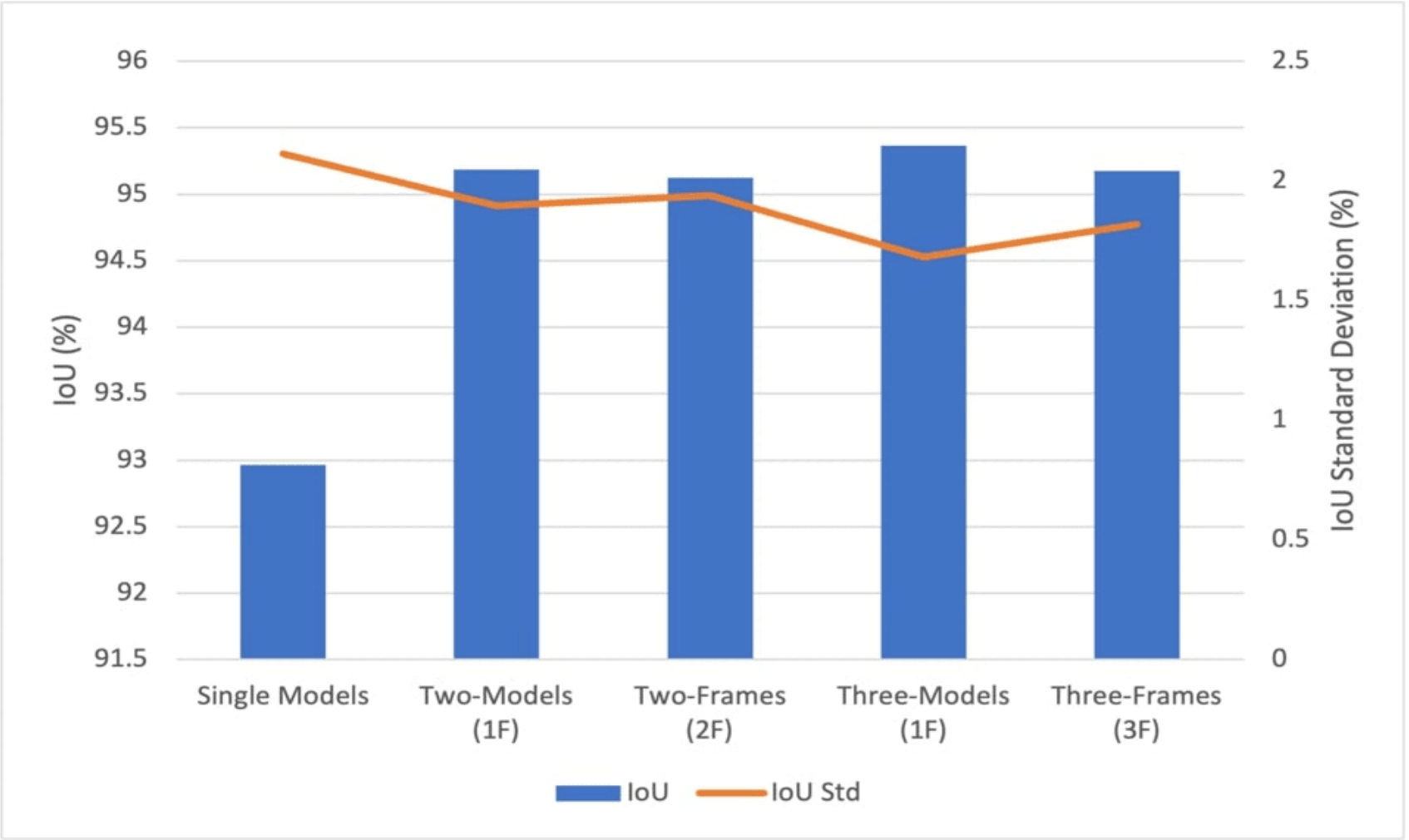
Proposed N-Frames Ensemble



Segmentation Models used for Experiment

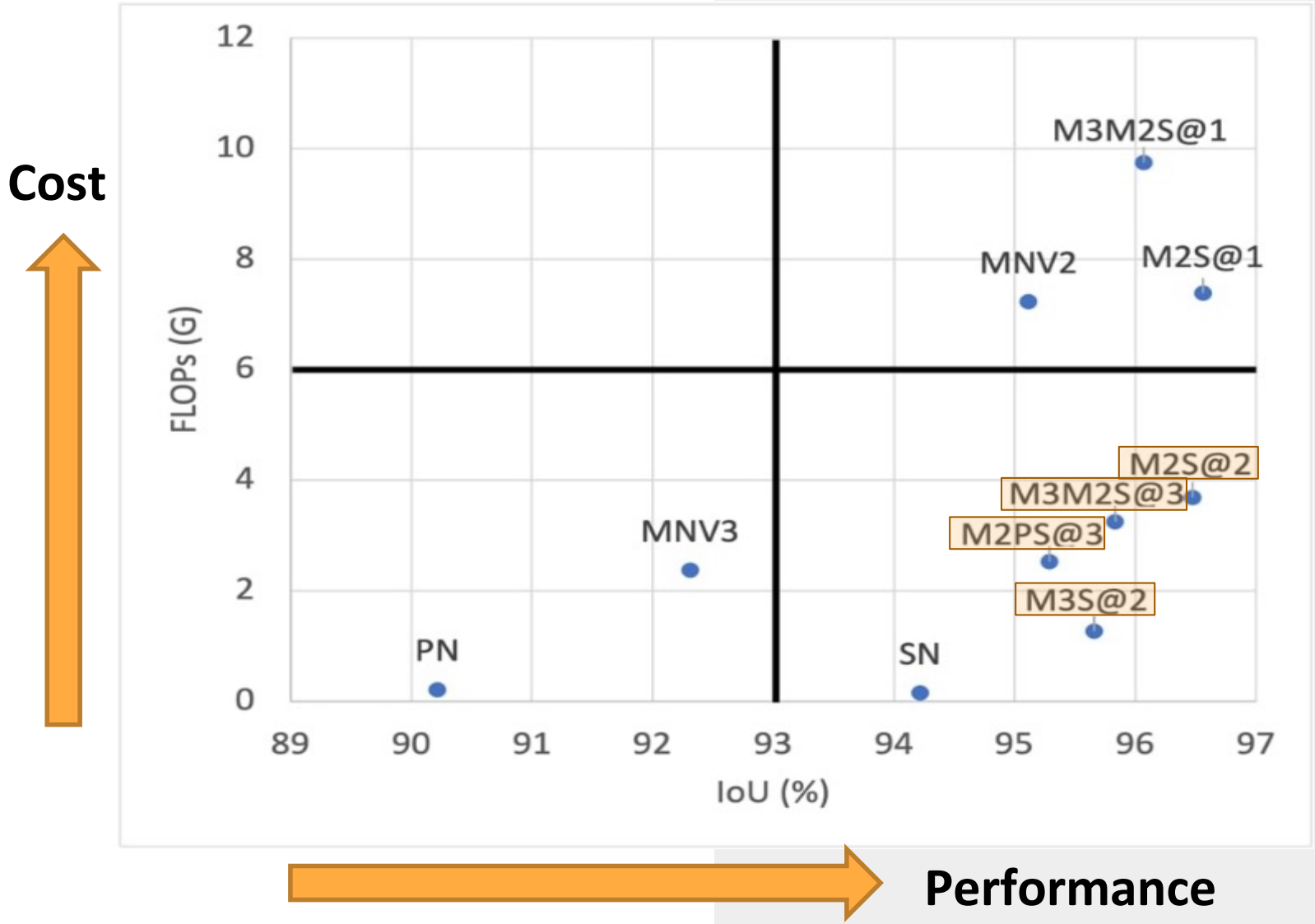
DSM	Backbone	IoU (%) of ref.	IoU (%) of our test
MNV3	MobileNetV3	94.19	94.76
MNV2	MobileNetV2	95.76	96.21
PN	MobileNetV2	95.99	95.7
SN	Custom	95.29	95.43

IoU (%) Comparison

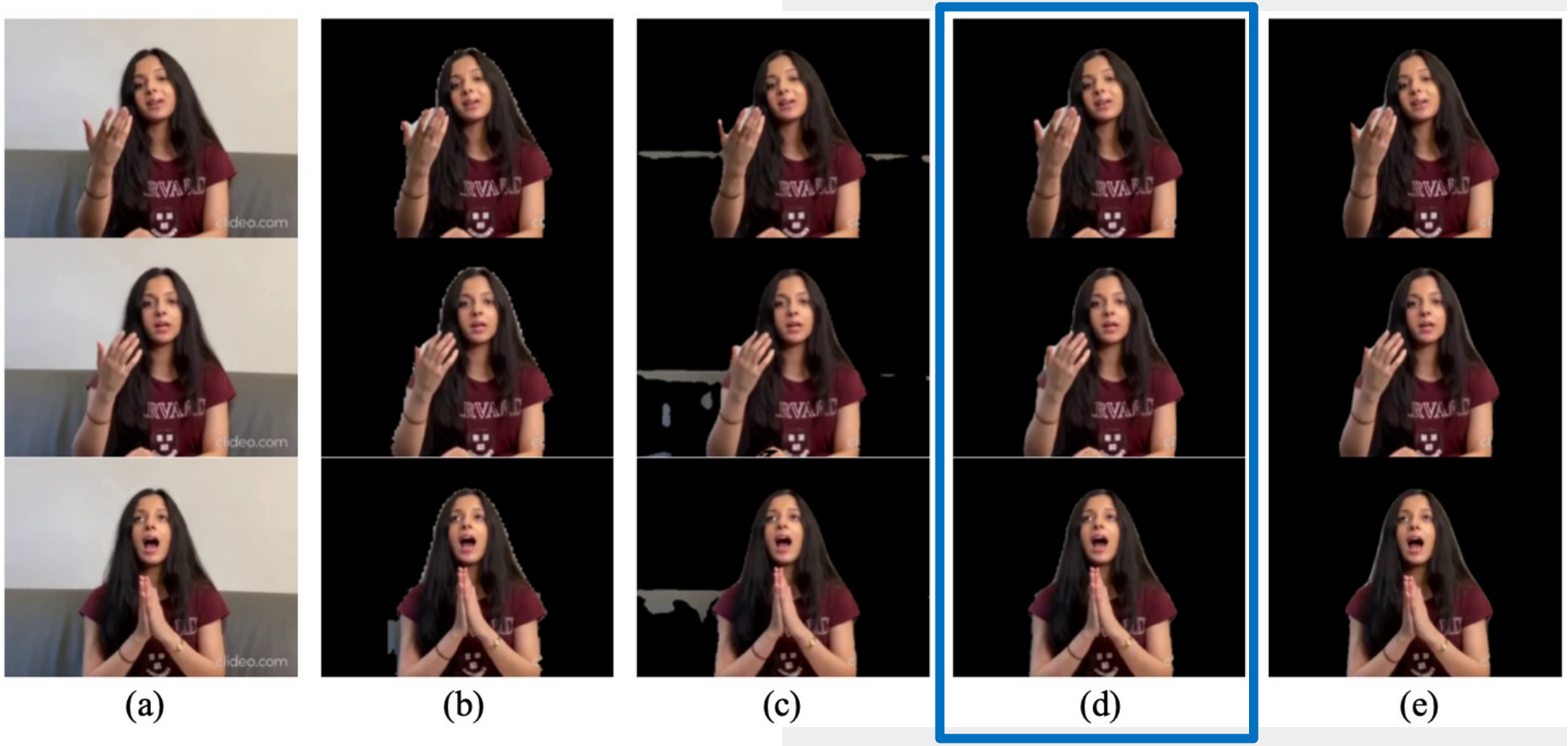


- IoU (Intersection over Union): A metric to measure the accuracy of segmentation.

Efficiency Comparison

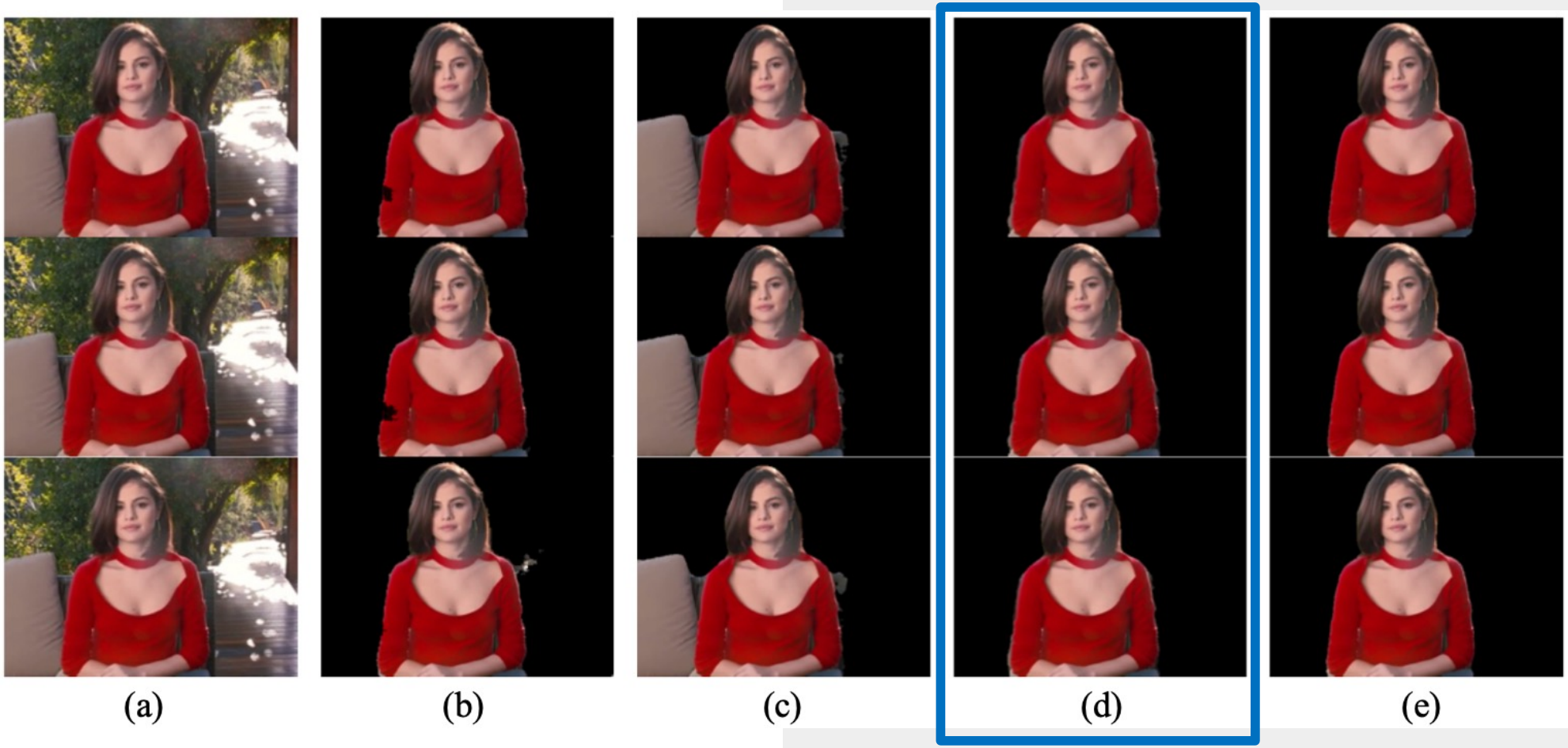


Segmentation result of different video frames



(a) Input frames; (b) output of Segmentation mode 1; (c) output of Segmentation model 2;
(d) output of proposed N-Frames Ensemble; (e) expected output (ground truth)

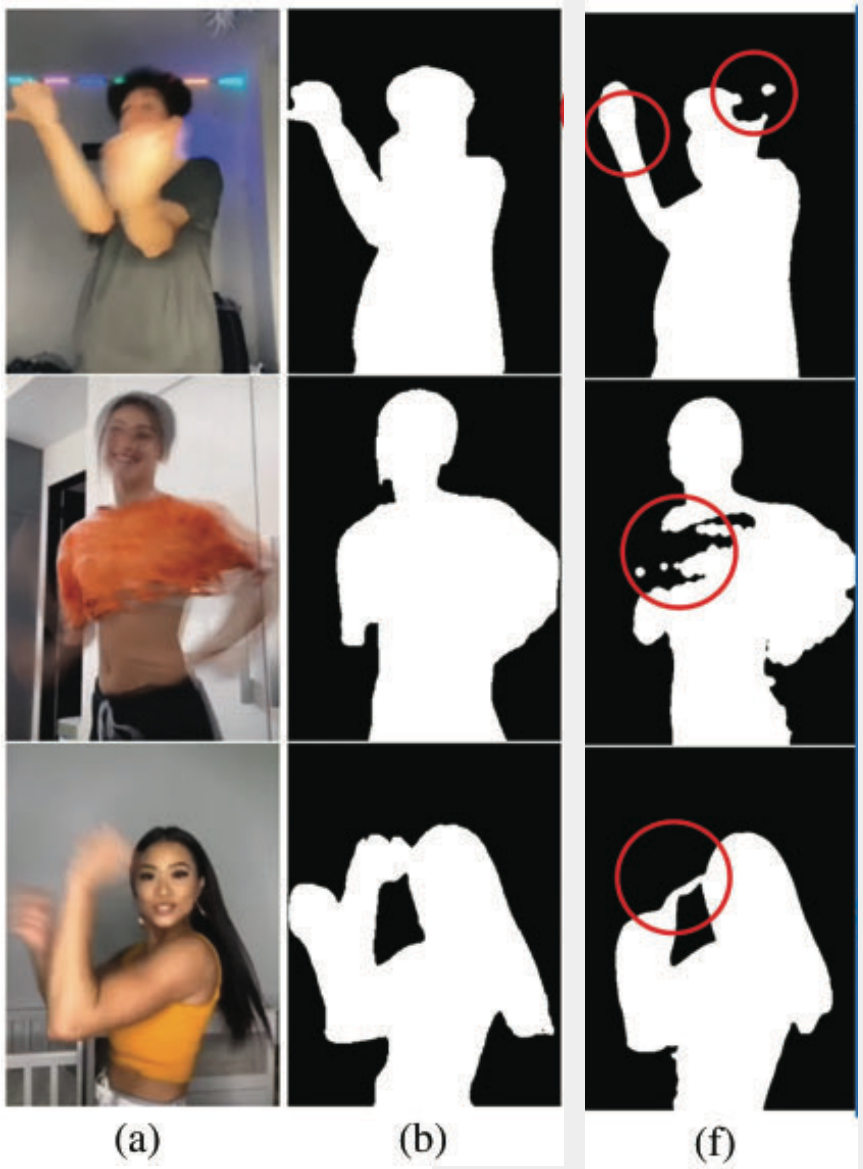
Segmentation result of consecutive video frames



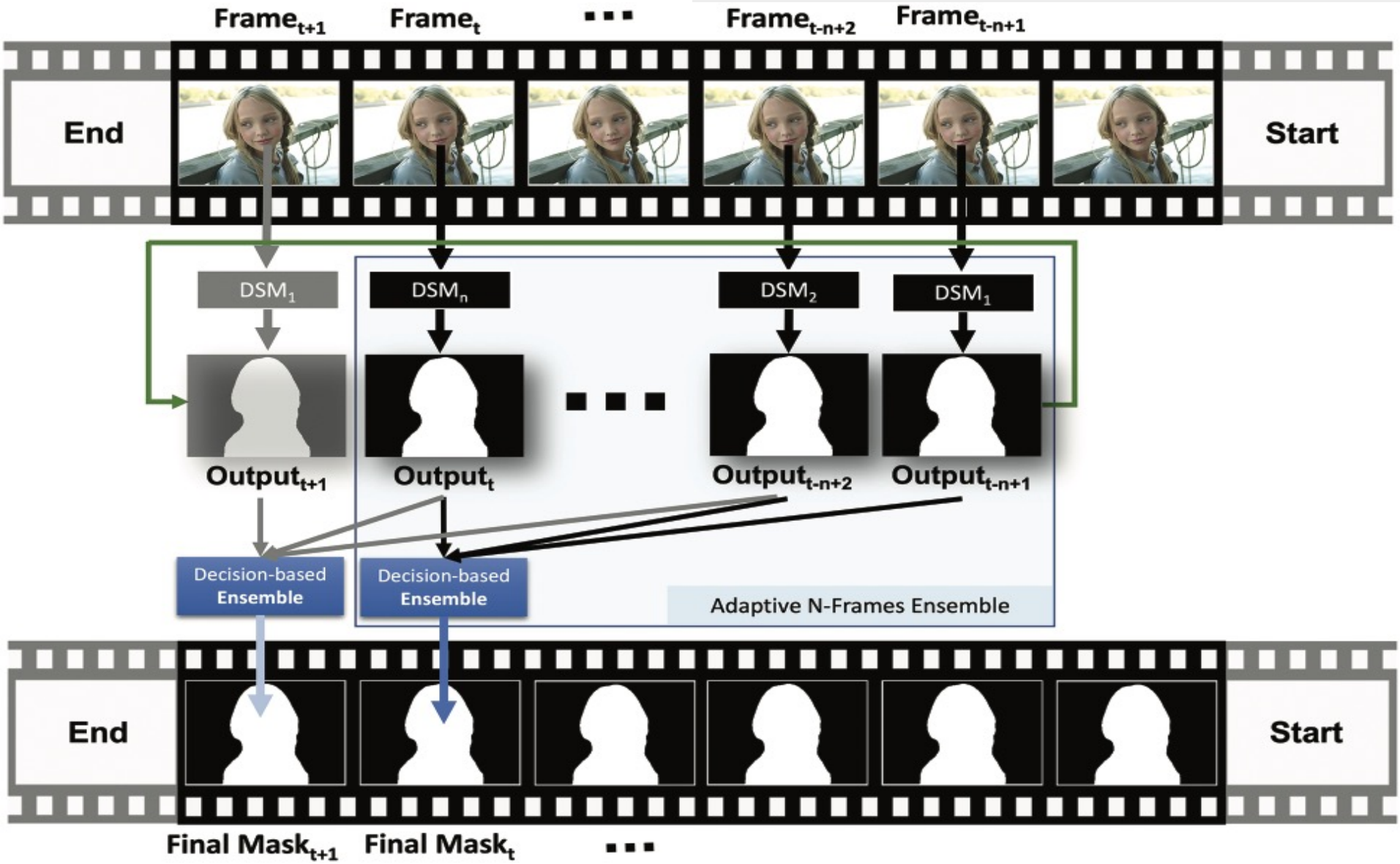
(a) Input frames; (b) output of Segmentation mode 1; (c) output of Segmentation model 2;
(d) output of proposed N-Frames Ensemble; (e) expected output (ground truth)

[Approach-3]
PVS using Adaptive N-Frames Ensemble of DSMs

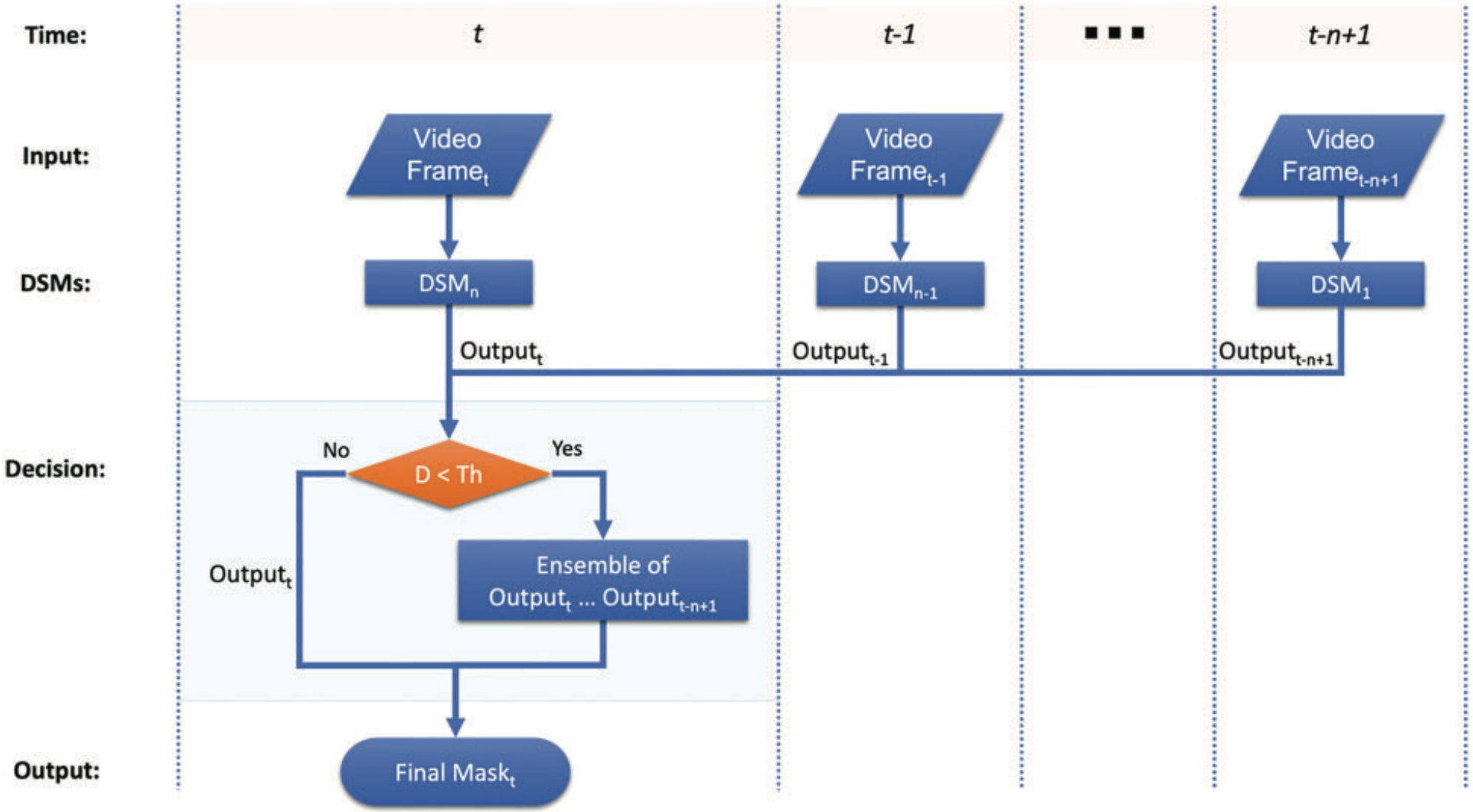
Limitation of N-Frames Ensemble (Fast moving object)



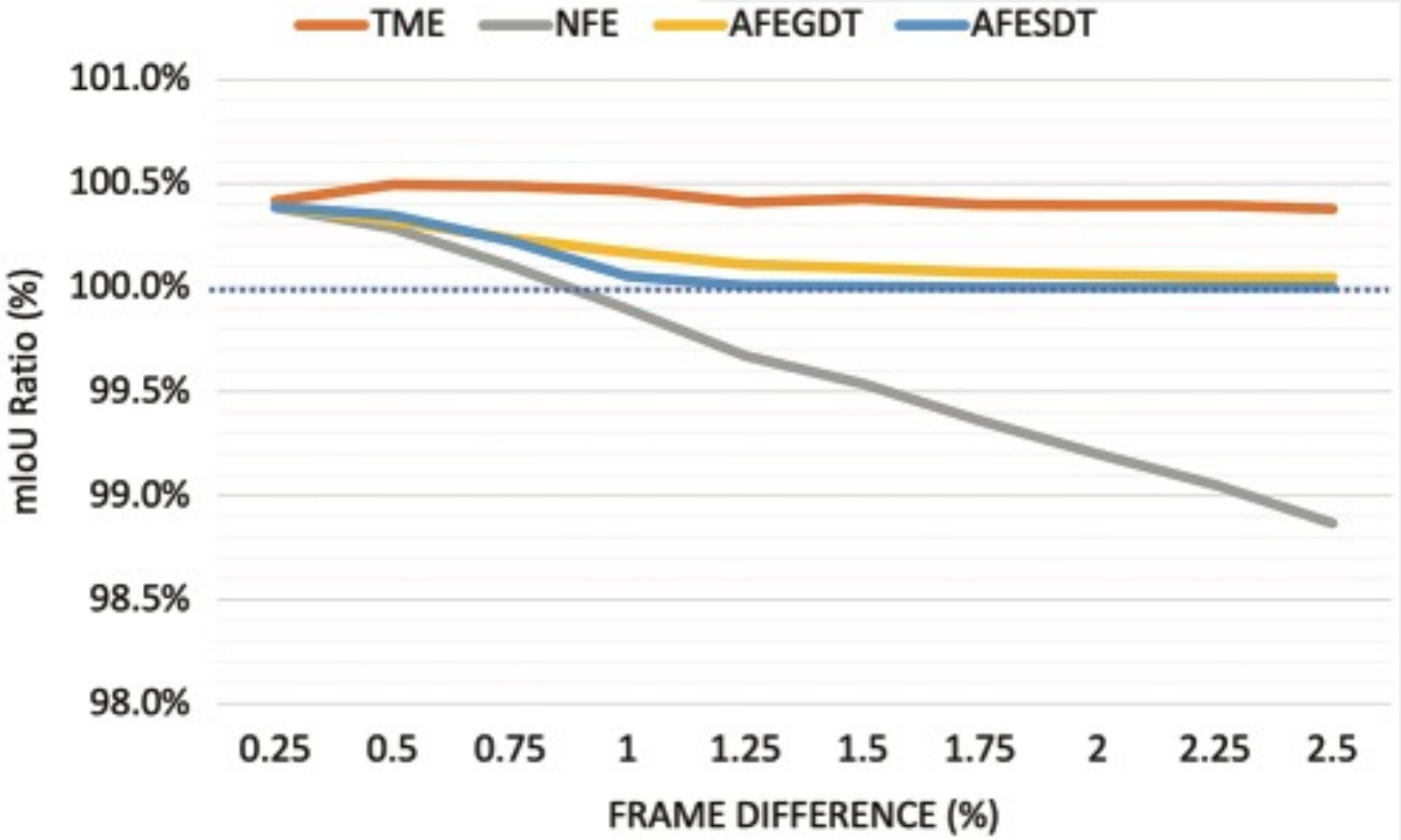
Proposed Adaptive N-Frames Ensemble



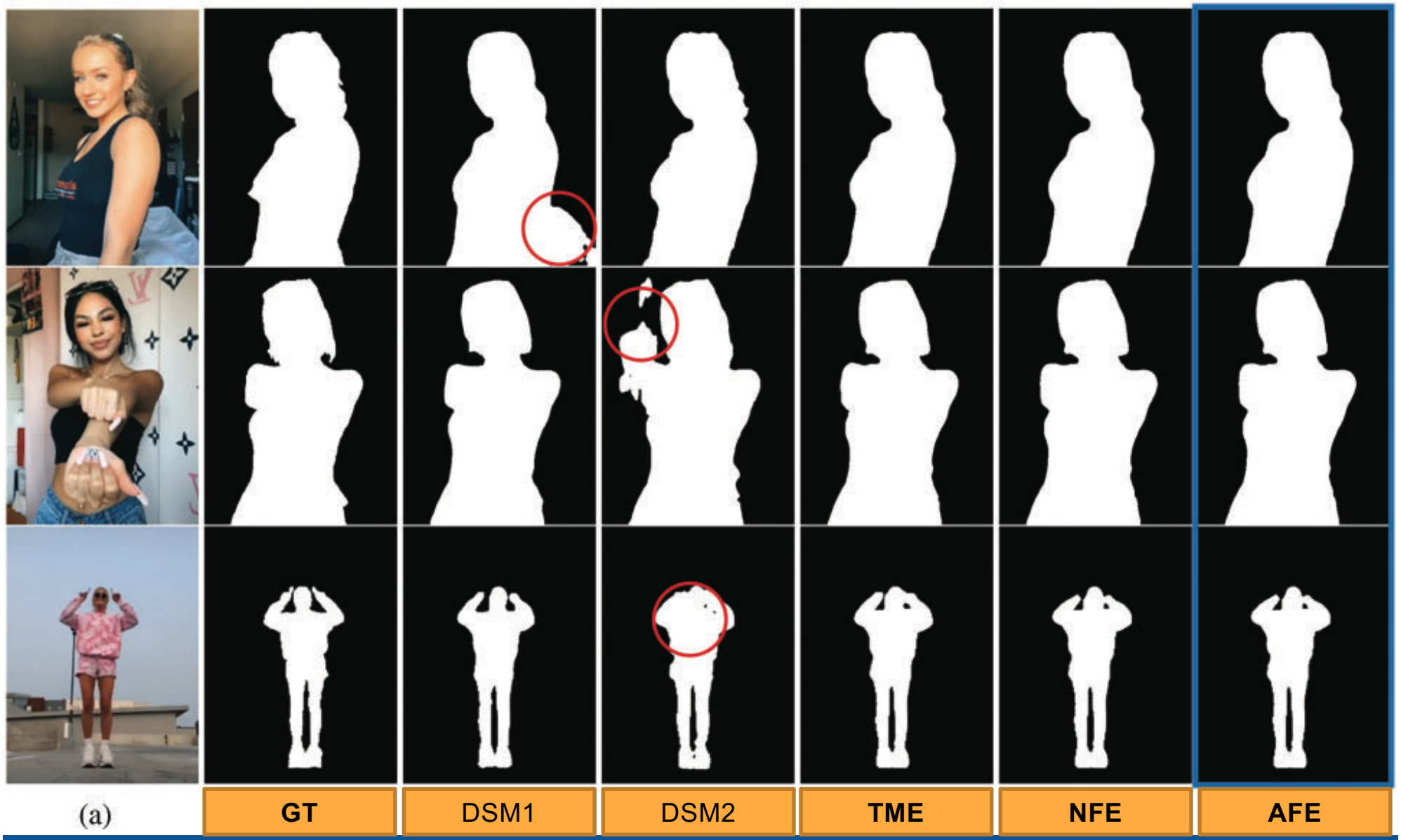
Decision-based Ensemble



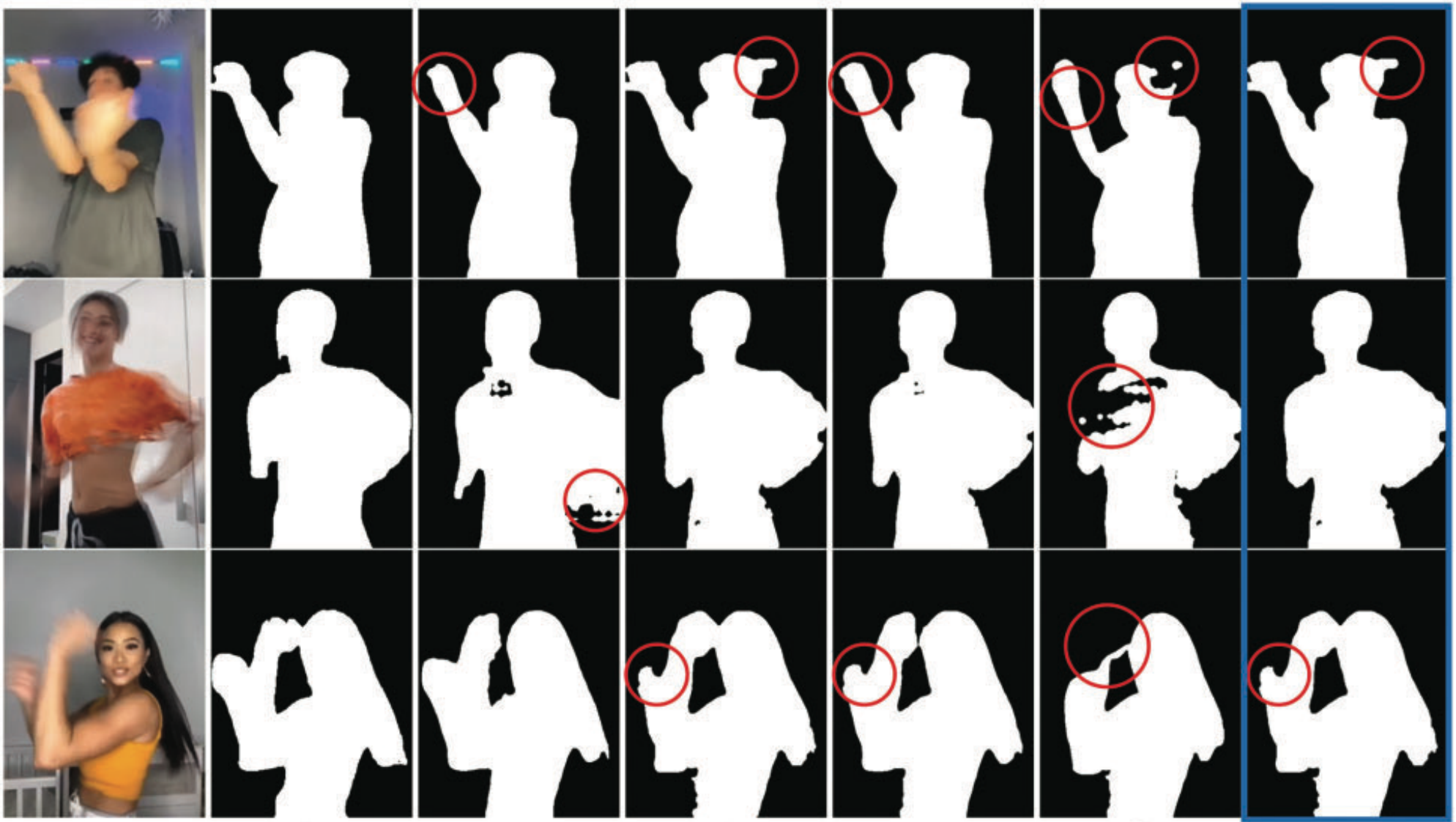
IoU COMPARISON BASED ON FRAME DIFFERENCE



Segmentation results of low movement



Segmentation results of high movement



Downscaling portrait images (common)

Comparison of SSIM and Execution Time

(SSIM: Structural Similarity Index Measure)

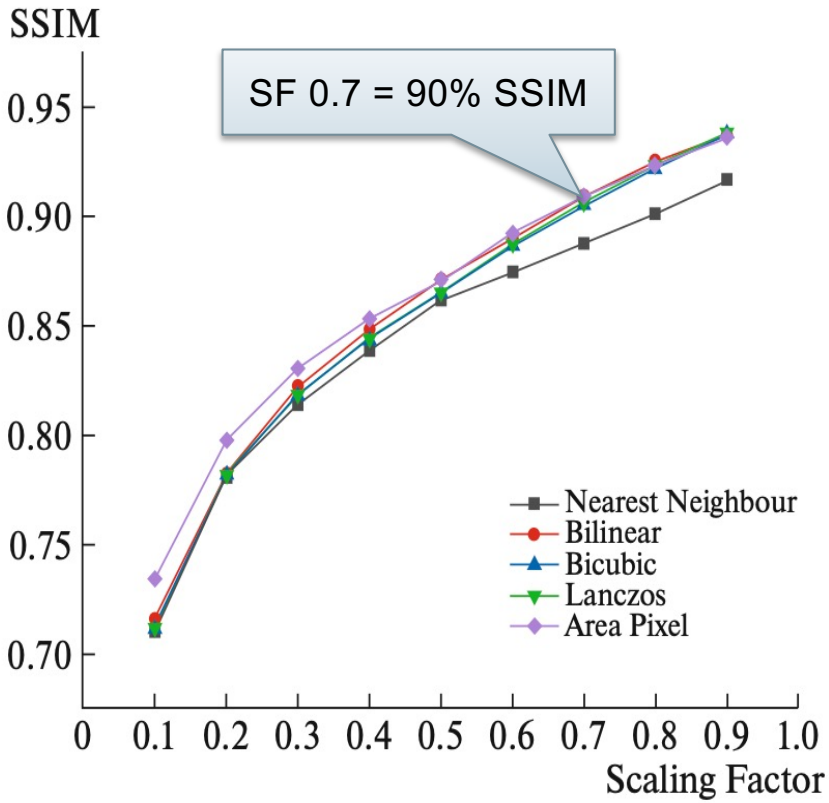


Fig. 2. SSIM values at different scaling factors for the downscaled images.

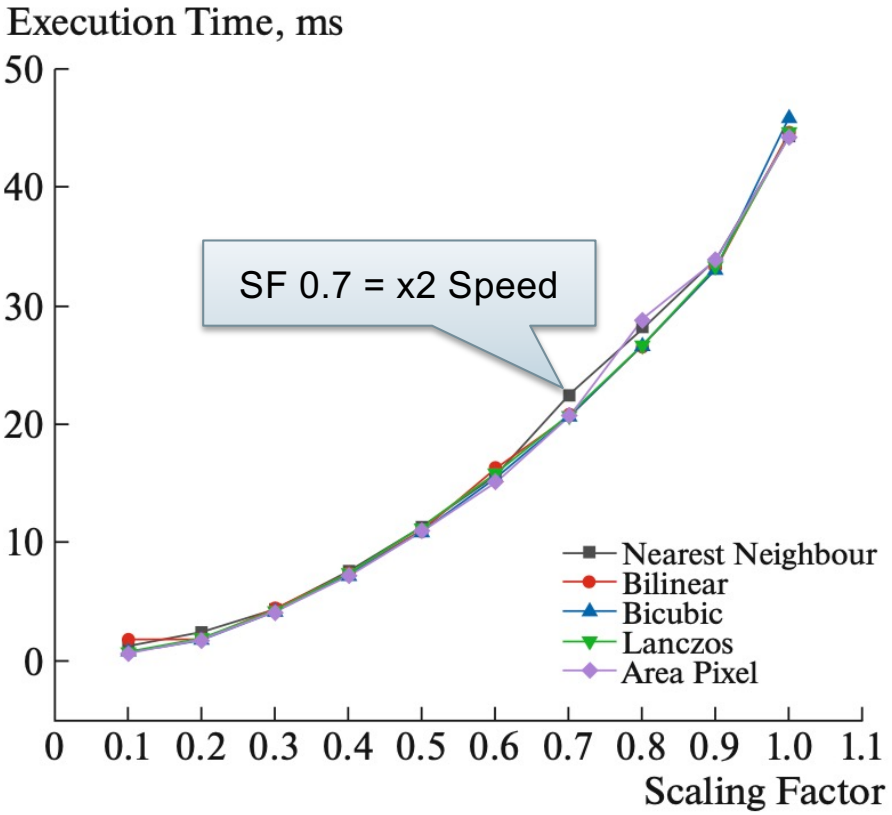


Fig. 5. Execution time of Canny Edge Detection for different scaling factor on the downscaled images.

Image size and speed comparison of DSMs

Table 5
Quantitative performance comparison. FLOPs are estimated with the size in brackets.

Method	FLOPS (G)	Parameters (M)
PortraitFCN + ($6 \times 224 \times 224$)	62.89	134.27
ENet($3 \times 224 \times 224$)	0.44	0.36
BiSeNet($3 \times 448 \times 448$)	9.52	12.4
BiSeNet($3 \times 224 \times 224$)	2.38	12.4
PortraitNet(ours, $3 \times 224 \times 224$)	0.51	2.1

(Source: “[2019] PortraitNet - Real-time portrait segmentation network for mobile device”)

Image size and speed comparison of DSMs

N	Backbone	RF2	SH	F	mIOU	Params	MAdds	CPU (f)	CPU (h)
1	V2	-	×	256	72.84	2.11M	21.29B	3.90s	1.02s
2	V2	✓	×	256	72.56	1.15M	13.68B	3.03s	793ms
3	V2	✓	✓	256	72.97	1.02M	12.83B	2.98s	786ms
4	V2	✓	✓	128	72.74	0.98M	12.57B	2.89s	766ms
5	V3	-	×	256	72.64	3.60M	18.43B	3.55s	906ms
6	V3	✓	×	256	71.91	1.76M	11.24B	2.60s	668ms
7	V3	✓	✓	256	72.37	1.63M	10.33B	2.55s	659ms
8	V3	✓	✓	128	72.36	1.51M	9.74B	2.47s	657ms
9	V2 0.5	✓	✓	128	68.57	0.28M	4.00B	1.59s	415ms
10	V2 0.35	✓	✓	128	66.83	0.16M	2.54B	1.27s	354ms
11	V3-Small	✓	✓	128	68.38	0.47M	2.90B	1.21s	327ms

(Source: “[2019] Searching for MobileNetV3 ”)

Summary of Approach-1

Objective-1		Approach	Contribution & Result
PS for selfie photos using ensemble method	**Speed**	Optimal combination of DSMs.	The optimal combinations of DSMs were explored and experimented.
			Three models ensemble (MNV3+PN+SN) showed 96.5% accuracy using Giga 2.7 Floating Point Operations (FLOPs) computing power while MNV2 showed 96.2% accuracy using 7.2 Giga FLOPs.
	Accuracy	Ensemble of heterogeneous DSMs	A novel approach using the ensemble of heterogeneous DSMs for portrait image segmentation was proposed and evaluated using the EG1800 dataset.
			The proposed method was superior to single DSMs in terms of accuracy, bias error and variance error. It showed more than 1.4% higher accuracy than single DSMs on average.

Summary of Approach-2

Objective-2		Approach	Contribution & Result
PS for portrait videos using ensemble method	Speed	Rotation of multiple DSMs.	A novel approach using the rotation method of DSMs for portrait video segmentation was introduced and experimented.
			The required FLOPs of PS processing was 2.49 Giga FLOPs for both the proposed method and single DSMs on average.
	Accuracy	N-Frames ensemble of DSMs.	A novel N-Frames ensemble of DSMs for portrait video segmentation was proposed and evaluated.
			The proposed approach performed with more than 2% higher accuracy than single DSMs on average. It also showed lower variance and bias error than single DSMs.

CONCLUSION & FUTURE WORK

- The applications using PS technology are increasing more and more.
- Several studies on PS using single deep-learning model and object segmentation using the ensemble of DSMs were reported from 2016.
- However, the PS using ensemble approach is an area that needs further study.
- This work introduces three ensemble approaches of multiple DSMs for high-performance PS.
- The first ensemble approach for portrait image segmentation → more than 1.4% higher accuracy than sing DSMs.
- The second ensemble approach for portrait video segmentation → more than 2% higher accuracy than single DSMs while consuming equal computing power to single DSMs.
- Future work is to apply the proposed models to the segmentation of other objects.

Appendix: Datasets

		Related works	This work	Note
General purpose (10 papers)	**Dataset**	Small number of images (2~11 images)	800 portrait images	**- No standard dataset.**
	Key point	Information preservation or efficiency	Information preservation + speed improvement	**- Our novelty.**
Deep-Learning (11 papers since 2016)	**Dataset**	EG1800(5), Own(3), PFCN(2), Pascal VOC(2), COCO(2), OCHuman(1), Baidu(1), Supervise.ly(1)	EG1800, Own	**- EG1800** is the most popular. **- Own dataset** is also used frequently.
	Key point	Downscaled image for high-speed	Downscaled image for high-speed	**- Downscaling is a common for speed improvement.**

Thank You