Self-adversarial Multi-scale Contrastive Learning for Semantic Segmentation of Thermal Facial Images

Jitesh Joshi, Nadia Bianchi-Berthouze, Youngjun Cho*

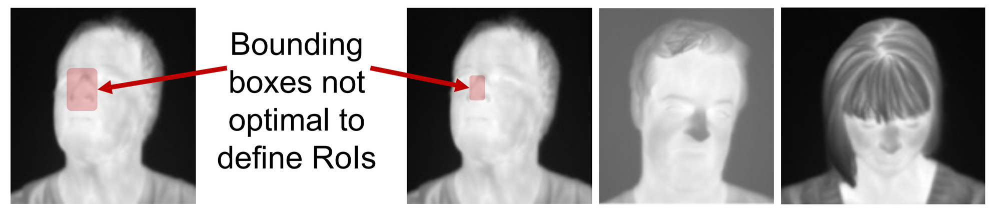
Department of Computer Science, University College London, UK

UCL

Motivation: Challenges in Identifying Regions of Interest

Thermal facial images exhibit:

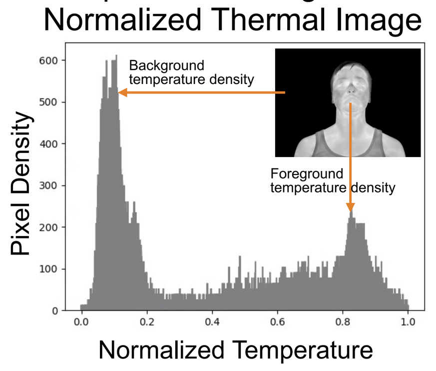
- Less salient facial features.
- Physiology driven changing spatio-temporal patterns.
- Pose variations.
- Occlusions such as hairs, eyeglasses.
- Changing spatio-temporal patterns due to varying ambient thermal conditions.



- RoIs need pixel-wise anatomical definition[1].

Semantic Segmentation

Temperature Histogram of Normalized Thermal Image



- Typical thermal image in controlled settings shows higher foreground temperature (e.g., facial region) compared to the background.
- This characteristic can potentially overfit the segmentation model, with potential failure when this condition is violated, e.g., presence of hot beverages in the scene.

1.A. Duarte et al., 'Segmentation Algorithms for Thermal Images', Procedia Technology, vol. 16, pp. 1560–1569, Jan. 2014, doi: 10.1016/j.protcy.2014.10.178.
2.Kopaczka et al., 'A fully annotated thermal face database and its application for thermal facial expression recognition'. I2MTC Conference, 2018. doi: 10.1109/I2MTC.2018.8409768. ZSCC: 0000021

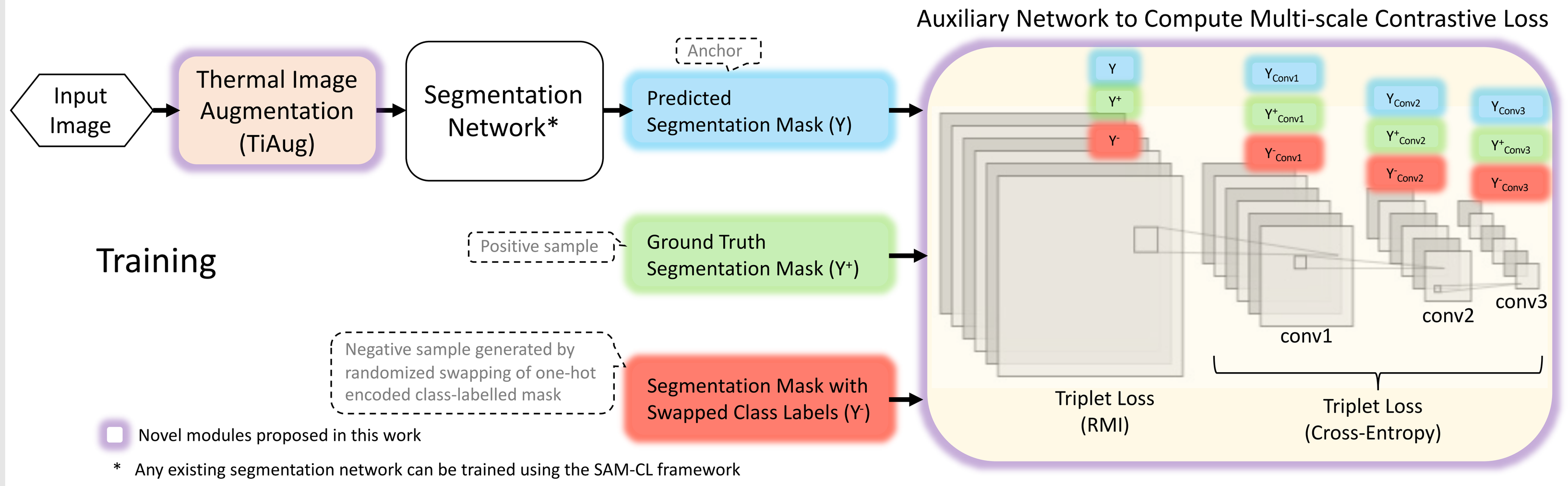
Proposed Framework: Self-adversarial Multi-scale Contrastive Learning (SAM-CL)

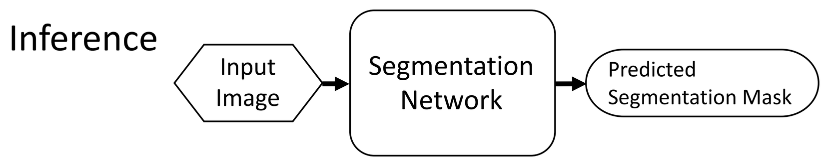
This work introduces SAM-CL framework to address below challenges in training existing segmentation networks for thermal images:

- Limited size of datasets.
- Unavailability of thermal images acquired in real-world settings such as:
- Unavailability of pretrained weights that are essential for applying existing state-of-the-art (SOTA) learning strategies, such as supervised contrastive learning.

SAM-CL framework consists of:

- **Thermal image augmentation (TiAug) module** to generate representations of unconstrained thermal settings by applying domain-specific transformations to the thermal images acquired in controlled settings.
- **Self-adversarial multi-scale contrastive loss function**, that utilizes the adversarial attacks made by the TiAug module to achieve intra-class proximity and inter-class separation of the feature representations.



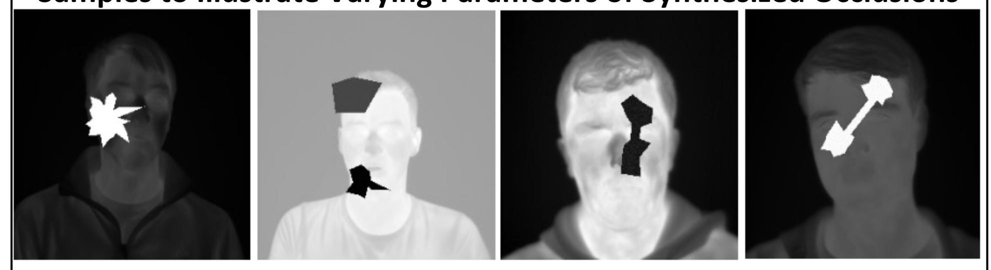


SAM-CL framework is needed only for training the segmentation network, resulting in no computational overhead during inference.

Proposed Thermal Image Augmentation (TiAug) Module

Thermal Occlusions f_{occ}	Thermal Noise η^{HxW}	Basic Image Manipulation
Synthesized occlusions to simulate real-world objects	Pixel-wise noise (max = 0.1 °C; ~ NETD of mobile thermal cameras)	Geometric Transformations

Samples to Illustrate Varying Parameters of Synthesized Occlusions



Single occlusion; Temperature higher than foreground (facial regions)	Two occlusions; Both at different and lower temperature than foreground	Two connected occlusions; Both at same and lower temperature than foreground	Two connected occlusions; Both at same and higher temperature than foreground

$$I_{aug}^{HxW} = f_{occ}(I_{org}^{HxW}, g(\vartheta_{sz}, \vartheta_{sh}, \vartheta_{temp}, \vartheta_{xy}, \vartheta_{config})) + \eta^{HxW}$$

Parameters for Synthesized Occlusions

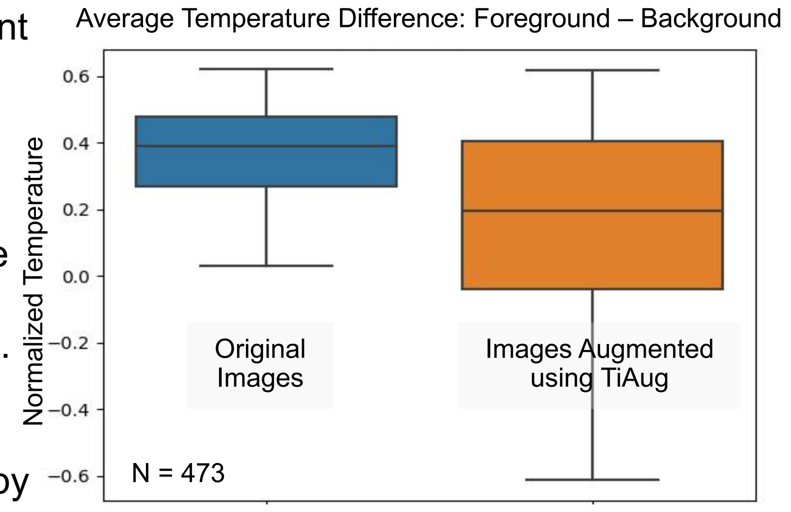
- Size: ~1% to ~40% of face area
- Shape: Irregularity – 0 to 10%; Vertices – 2 to 30; Spikeness – 0 to 50%
- Temperature: Min – (Max – Min); Max + (Max – Min)
- Position: Across entire image
- Configurations: Single; Dual; Connected Objects

Submodules proposed in this work

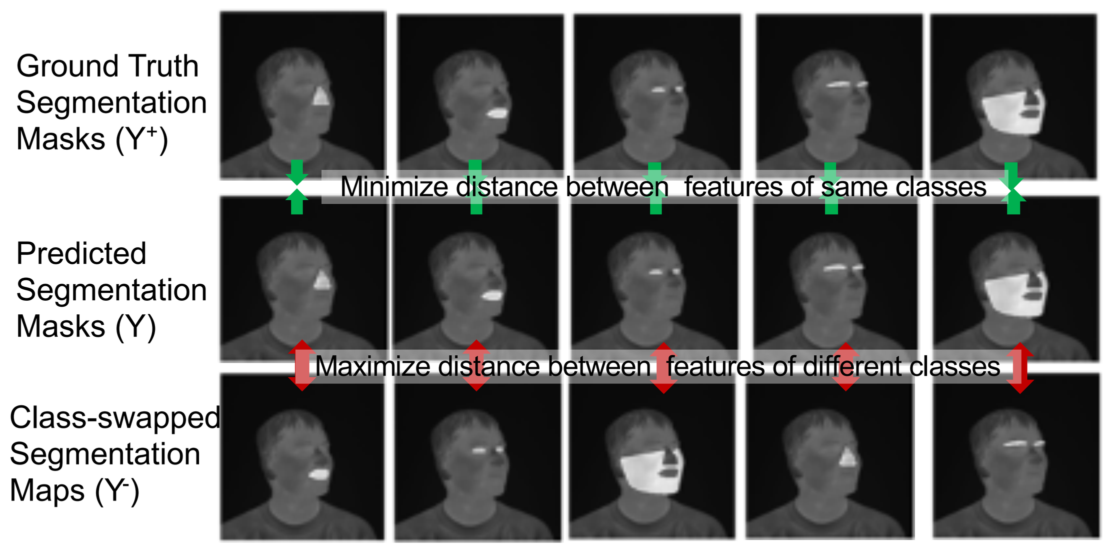
Effectiveness of the TiAug Module

- Typically, the foreground (facial region) is at higher temperature than background or ambient temperature, when images are acquired in controlled settings.
- As evidenced by the box-plot analysis, the TiAug module is effective in generating diverse foreground-background temperature distributions, that represent real-world settings.
- The TiAug module presents self-adversarial attacks for training the segmentation network by generating challenging real-world appearances.

Average Temperature Difference: Foreground – Background



Proposed SAM-CL Loss Function



- Predicted segmentation mask (Y) or logits, ground-truth mask (Y⁺) and class-swapped mask (Y⁻) are passed through the auxiliary network and down-convolved.
- Triplet loss is applied at each layer, resulting in multi-scale contrastive loss function.

$$\mathcal{L}_{s0}(Y_{oh}, Y_{oh}^{+}, Y_{oh}^{-}) = \max\{d(Y_{oh}, Y_{oh}^{+}) - d(Y_{oh}, Y_{oh}^{-}) + margin, 0\}$$

$$\mathcal{L}_{SAM-CL} = \mathcal{L}_{s0}(Y_{oh}, Y_{oh}^{+}, Y_{oh}^{-}) + \mathcal{L}_{s1}(Y_{Conv1}, Y_{Conv1}^{+}, Y_{Conv1}^{-}) + \mathcal{L}_{s2}(Y_{Conv2}, Y_{Conv2}^{+}, Y_{Conv2}^{-}) + \mathcal{L}_{s3}(Y_{Conv3}, Y_{Conv3}^{+}, Y_{Conv3}^{-})$$

Quantitative Performance Analysis

Segmentation Network	Learning Strategy (Loss Function)	mIoU (%)
UNET [152]	Pixel-wise Segmentation (BCE)	67.64
	Pixel-wise Segmentation (DICE)	75.00
	GAN (SegAN) [190]	76.79
	GAN (SegGAN) [200]	75.50
	RMI [201]	81.35
	ContrastiveSeg [183]	81.24
	SAM-CL (Ours)	**82.11 (+0.76)**
DeepLabV3+ResNet101 [29, 66]	RMI [201]	75.85
	ContrastiveSeg [183]	74.45
	SAM-CL (Ours)	**79.29 (+3.44)**

Segmentation Network	Learning Strategy (Loss Function)	mIoU (%)
Attention UNET [138]	Pixel-wise Segmentation (BCE)	66.61
	Pixel-wise Segmentation (DICE)	75.14
	GAN (SegAN) [190]	76.75
	GAN (SegGAN) [200]	76.24
	RMI [201]	81.39
	ContrastiveSeg [183]	81.50
	SAM-CL (Ours)	**82.85 (+1.35)**
HRNetV2-W48 [171]	RMI [201]	78.46
	ContrastiveSeg [183]	78.36
	SAM-CL (Ours)	**78.97 (+0.61)**

Ablation Study

Segmentation Network	mIoU (%) Performance		
	RMI	RMI + TiAug	RMI + TiAug + SAM-CL
UNET [152]	81.36	81.91	**82.11**
Attention UNET [138]	81.39	82.29	**82.85**
HRNetV2-W48 [171]	78.13	78.87	**78.97**
DeepLabV3+ResNet101 [29, 66]	75.85	78.07	**78.12**
DeepLabV3+Xception [29, 41]	76.55	77.31	**77.85**

- Consistent performance gains are observed for all the segmentation networks when thermal images are augmented using the TiAug module and the SAM-CL loss function is used for optimization.

Qualitative Performance Analysis