BOUNDARY GUIDED IMAGE TRANSLATION FOR POSE ESTIMATION FROM ULTRA-LOW RESOLUTION THERMAL SENSOR

Kohei Kurihara†, Tianren Wang ‡, Teng Zhang ‡, Brian Lovell ‡

†Advanced Technology R&D Center Mitsubishi Electric Corporation, ‡School of ITEE, The University of Queensland, Australia

MOTIVATION

Pose estimation using low resolution thermal sensor

✓ Capable to be installed in home appliance in low cost
✓ Capable to acquire in low/no light condition
✓ Not invade people's privacy
X Pose estimation(PE) task is challenging
- large domain gap between the thermal and visible
- low resolution, lack of texture and boundary

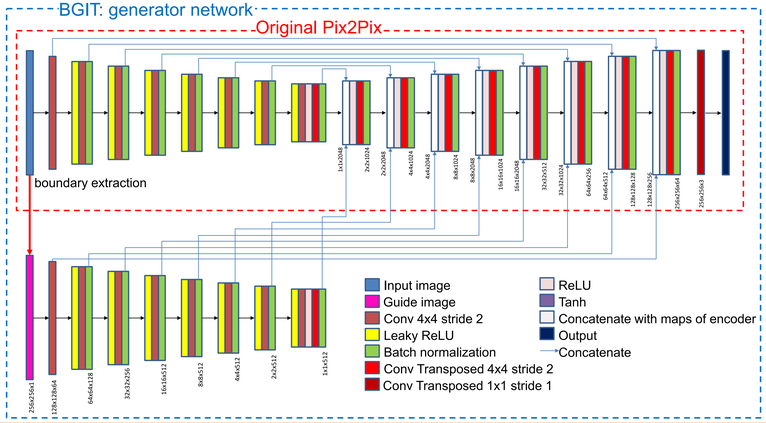
Solution

- Utilize image to image (I2I) translation technique
  Using I2I as a pre-processing step, we can utilize off-the-shelf SOTA visible light PE models without the hassle of re-training the models.
- Develop a new I2I translation architecture which can translate the original blurred thermal image into a visible light image **with sharper boundaries.**

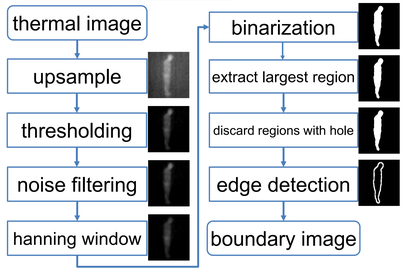
METHODS

BGIT: Boundary Guided I2I Translation

- Building our architecture based on Pix2Pix, adding boundary feature map module
- Our network encoded boundary features and simply concatenated them to vectors of thermal images



Boundary Extraction



Example of an extracted boundary

EXPERIMENT RESULTS

Task A. Pose Estimation result

TABLE I.
RESULT OF PCK@0.2 (IN % HIGHER IS BETTER)

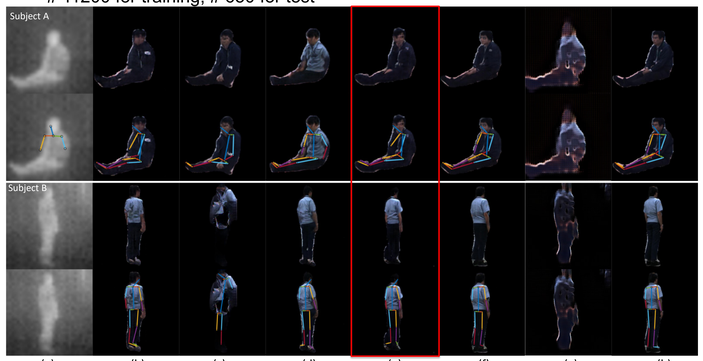
Method/	PCK@0.2	difference from (b)
(a) input thermal	8.9	-46.570
(b) Pix2Pix-T[1]	55.5	0.000
(c) Pix2Pix-C[1]	40.0	-15.470
(d) BGIT-T+C	59.9	4.430
(e) BGIT-T+B	**63.6**	**8.080**
(g) DRIT[2]	0.04	-55.430
(f) BGIT-T+VB (upper limit)	73.9	18.390

T:thermal image
C:canny edge
B: boundary from thermal
VB: boundary from visual

Pose estimator: OpenPose

PCK: Percentage of Correct Key-points (Distance between predicted and true joint < 0.2 torso diameter)

■Dataset: Thermal human pose dataset(THP dataset)
- Thermal : 32 × 32 pixels→ 256 ×256 pixels (upsample)
- Visual : 256 ×256 pixels
  # 11200 for training, # 650 for test



Task B. Image to Image translation result

TABLE II
RESULT OF AVERAGE PSNR AND SSIM, HIGHER IS BETTER

Method	PSNR	SSIM
(b) Pix2Pix-T[1]	**15.80**	0.531
(c) BGIT-T+C	15.56	0.525
(d) BGIT-T+VC	15.43	**0.538**
(e) DRIT[2]	13.27	0.455
(f) CannyGAN[3]	12.9	0.457

Task B is to demonstrate the superior image quality from BGIT I2I architecture

T: thermal image
C: canny edge from thermal
VC: canny edge from visual

■Dataset: FLIR thermal dataset(THP dataset)
- Thermal : 32 × 32 pixels→ 256 ×256 pixels (upsample)
- Visual : 256 ×256 pixels
  # 8300 for training, # 1200 for test

CONCLUSION

I. We aim to address the pose estimation task in ultra-low resolution thermal images without labeling and re-training/finetuning
II. We proposed a new I2I translation architecture that uses boundary information as a guide in addition to the original thermal image
III. Analysis reveals that the proposed method is able to generate images of higher quality than other state-of-the-art methods

REFFERENCE

[1]P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In CVPR, 2017.
[2]Hsin-Ying Lee, Hung-Yu Tseng, Jia-Bin Huang, Maneesh Singh, and Ming-Hsuan Yang, "Diverse image-toimage translation via disentangled representations," In ECCV, 2018, pp. 35–51.
[3]Wang, Tianren, et al. "Cannygan: Edge-Preserving Image Translation with Disentangled Features." 2019 IEEE International Conference on Image Processing (ICIP). IEEE, 2019.