Supplementary Material

Supplementary Material for the paper "CompNet: Competitive Neural Network for Palmprint Recognition Using Learnable Gabor Kernels," DOI: 10.1109/LSP.2021.3103475.

A. Parameter Selection

1) Experimental Settings: The experimental settings are same as that used in the paper. In all the experiments presented in this work, ROI images were resized to 128×128 if they were not of this size. Besides, the color images were converted to grayscale images before use. According to our observation, ID-121 and ID-122 in the REST dataset are the same; thus, ID-122 is removed in our experiments. As a result, the REST dataset we used only contains 356 classes. For the IITD, IF, and IN datasets, the official ROI images were used. The initial learning rate of the CompNet was set to 0.001, which decreased by 20% every 300 epochs. All the CNN-based models were trained for 3000 epochs. Their batch sizes were set to eight, and the Adam optimizer was used.

2) Number of Directions: Table I shows the rank-1 accuracy and EER on different datasets obtained using the CompNet with different number of orientations. The best d values for different datasets are not the same. For example, when $d = 15$, the CompNet reached the highest rank-1 (99.85%) on the IF dataset; when $d = 9$, it reached the highest rank-1 accuracies on the IITD (98.77%) and REST (92.51%) datasets. However, considering the trade-off between time cost and recognition accuracy, we fixed d to nine in all the experiments presented in the paper.

3) Angular margin: Table II shows the rank-1 and EER obtained by the CompNet with different m values on different datasets. Based on the results, the angular margin m was set to 0.5 in all the experiments reported in the paper ($\varepsilon = 30$). According to our experiments, a smaller m is helpful for the stability of model training; however, the corresponding rank-1 reduced slightly.

4) Details of Convolution: For the convolution stride p in the LGC layer, a smaller p value does not obviously increase the accuracy but does increase computational costs significantly. Thus, p was fixed to three for all CBs of the LGC layer in our experiments. For the PPU, to make full use of the SCC, the strides of the Conv1 and Conv2 layers were set to one. Zero paddings were configured during convolutions across the entire network.

5) PPU Configurations: We also tested different PPU configurations. The results in Table III illustrate that the optimal kernel sizes for the Conv1 and Conv2 layers are 5×5 and 1×1, respectively. When the Maxpool layer was omitted, the CompNet's rank-1 accuracy dropped from 92.51% to 89.06% and from 89.63% to 86.52% for the 5×5 and 1×1 Conv1 configurations, respectively. According to this, the first configuration in Table III is selected. As shown in Table III, without the PPU, the CompNet's accuracy is significantly reduced (from 92.51% to 87.10%). Consequently, the PPU module can significantly improve the recognition accuracy of the CompNet.

B. Visualization of CBs

1) Feature Maps: Fig. 1 shows some typical images of the Gabor filters and corresponding feature maps. As can be observed, firstly, the LGC layer decomposed the input image into responses of different directions. Then, the responses were

TABLE I: Rank-1 (%) and EER (%) on Various Datasets Obtained by the CompNet with Different Number of Orientations of the CBs.

Number of orientations (d)		3	6	9	12	15	18
Tongji	Rank-1	99.98	99.78	**100**	**100**	99.92	99.98
	EER	**0.011**	0.098	0.018	0.014	0.040	0.056
IITD	Rank-1	98.53	98.61	**98.77**	98.85	98.85	98.69
	EER	0.628	0.683	0.628	**0.518**	0.601	0.573
REST	Rank-1	91.13	91.48	**92.51**	91.36	92.17	92.17
	EER	3.187	**2.995**	3.211	3.149	3.111	3.149
IF	Rank-1	99.49	99.71	99.56	99.71	**99.85**	99.64
	EER	0.146	**0.031**	0.146	0.067	0.096	0.073
IN	Rank-1	99.57	99.71	99.64	**99.78**	99.71	99.57
	EER	0.238	0.145	0.193	**0.073**	0.155	0.218

TABLE II: Rank-1 (%) and EER (%) on Various Datasets Obtained by the CompNet with Different m values.

m		1.0	0.5	0.4	0.3	0.2	0.1	0.05
Tongji	Rank-1	100	**100**	100	100	100	100	99.98
	EER	**0.004**	0.018	0.010	0.005	0.008	0.018	0.040
IITD	Rank-1	**99.02**	98.77	98.69	98.61	**99.02**	98.36	98.28
	EER	0.737	0.628	**0.532**	0.655	0.764	0.956	1.076
REST	Rank-1	90.90	**92.51**	92.05	91.36	91.24	89.75	89.29
	EER	4.109	**3.211**	3.264	3.394	3.418	3.725	3.622
IF	Rank-1	99.42	**99.56**	99.34	99.42	99.13	99.27	99.05
	EER	0.364	**0.146**	0.146	0.170	0.217	0.219	0.583
IN	Rank-1	99.42	**99.64**	**99.64**	99.49	99.20	98.26	98.77
	EER	0.340	**0.193**	0.218	0.200	0.363	0.508	0.483

TABLE III: Rank-1 (%) and EER (%) on the REST Dataset Obtained by the CompNet with Different PPU Configurations.

Conv1	Maxpool	Conv2	Additional module[a]	EER	Rank-1
5×5	2×2	1×1	–	**3.211**	**92.51**
3×3	2×2	1×1	–	3.379	91.71
5×5	2×2	3×3	–	3.303	91.01
1×1	2×2	1×1	–	4.109	89.17
1×1	2×2	–	–	4.378	89.63
1×1	–	–	–	4.570	86.52
1×1	–	1×1	–	4.481	87.10
5×5	–	1×1	–	4.263	89.06
5×5	2×2	1×1	✓	4.109	87.44
–	–	–	–	4.782	87.10

[a] Additional module: ReLU + BN + Conv(1×1@12) + ReLU + BN

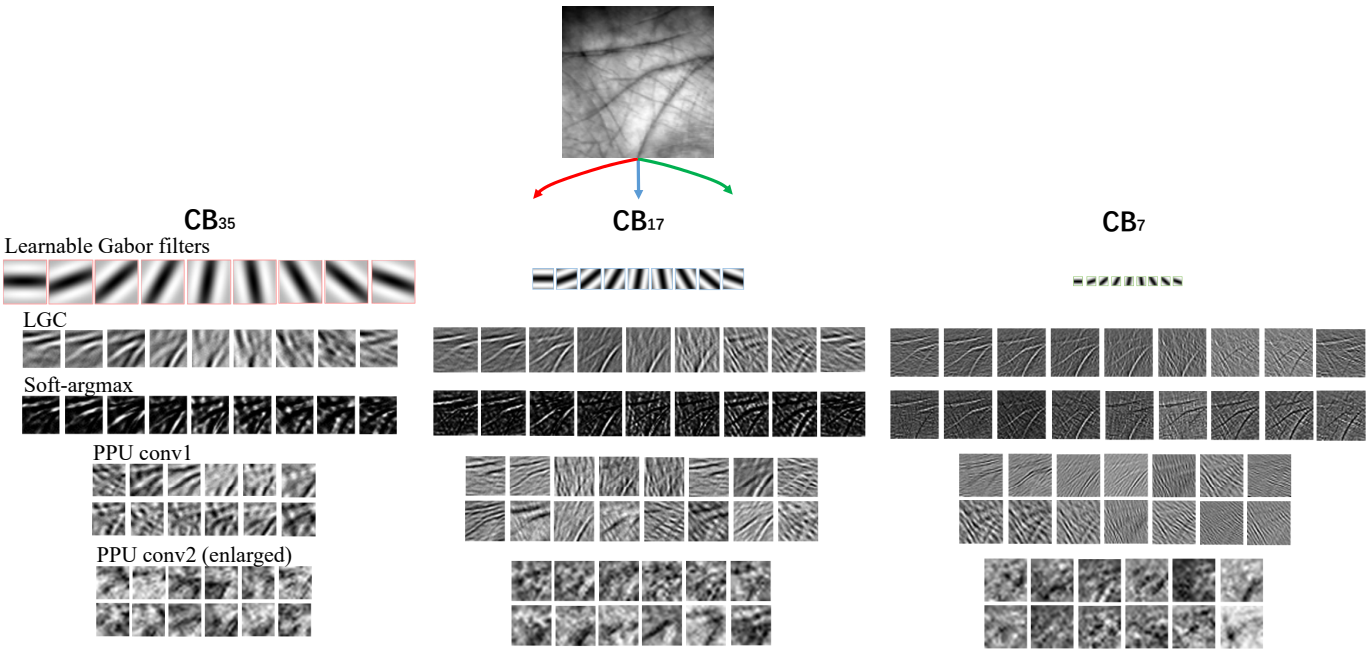


Fig. 1: Gabor kernels learned by the proposed CompNet and typical feature maps generated at different stages of an image from the Tongji dataset.

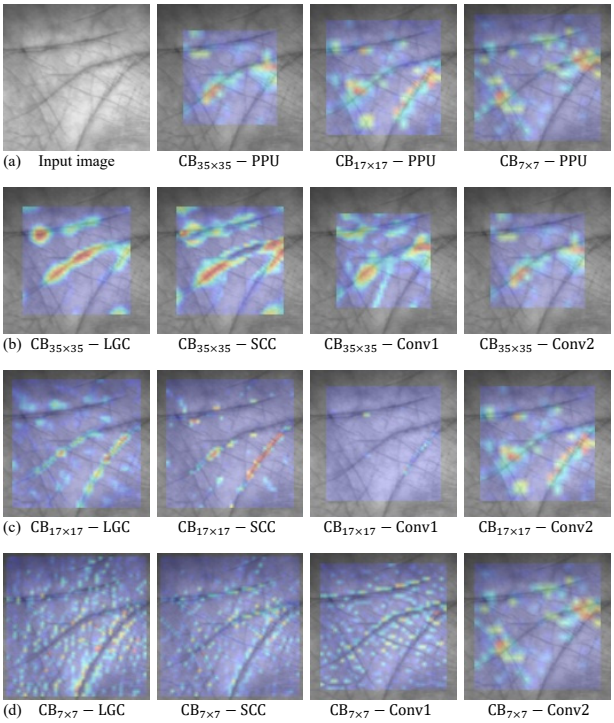


Fig. 2: Grad-CAM images obtained from different layers of the CompNet.

processed by channel-wise softmax operations. As a result, the feature maps became sparser than before, and the weak responses were suppressed. Finally, the PPU performed feature fusion and transformation to extract high-level knowledge.

2) Class Activation Mapping: The Grad-CAM technique [1] is utilized to observe the contributions of the SCC and PPU

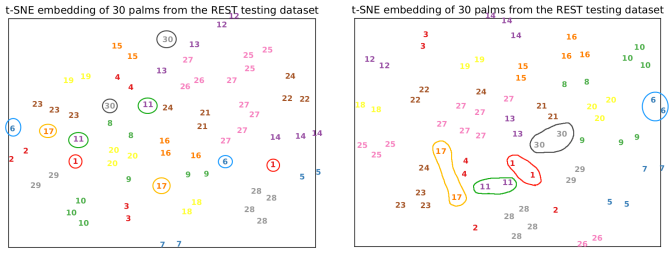
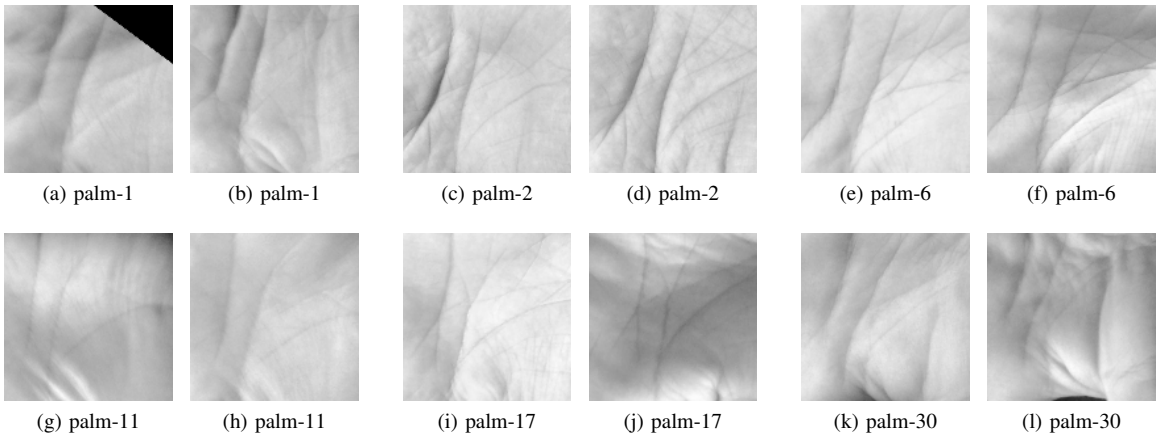


Fig. 3: Palm cluster distributions of features obtained from the FC layer via t-SNE. (a) CompNet without the PPU module, and (b) CompNet with the PPU module.

modules for classification. The heatmaps reflect which features are more critical for achieving higher discrimination. In Fig. 2, we can see that the PPU cares more about the palmprint line intersections, whereas the SCC only cares about the lines and edges, which are low-level features. It demonstrates that the PPU tried to extract high-level features by considering more spatial information. Thus, the PPU can increase the final accuracy.

3) Clustering Analysis: We utilized t-SNE [2] to do dimension reduction for the features extracted from the FC layer. Only two dimensions were reserved; thus, each palm feature is a point in the 2D feature space. The first 30 palms in the REST test set were selected for testing. Each palm has 2 to 5 images. To evaluate the contribution of the PPU (which is designed for further feature extraction), two configurations were tested in this experiment: one with the PPU, another one omitted the PPU. The palm distributions of the two configurations are shown in Fig. 3. From the figure, we can see that the PPU



(a) palm-1 (b) palm-1 (c) palm-2 (d) palm-2 (e) palm-6 (f) palm-6

(g) palm-11 (h) palm-11 (i) palm-17 (j) palm-17 (k) palm-30 (l) palm-30

Fig. 4: ROI images in the REST dataset.

is helpful to aggregate intra-class samples. The distributions of palm-1, palm-6, palm-11, palm-17, and palm-30 were improved. Corresponding ROI images are shown in Fig. 4. Figure 3 and 4 demonstrate that, when embedded with the PPU, the CompNet is more robust to lighting and localization variations. More *Supplementary Materials* are available at https://github.com/xuliangcs/compnet.

ACKNOWLEDGEMENT

Portions of the research in this paper use the REST'2016 Database collected by the Research Groups in Intelligent Machines, University of Sfax, Tunisia. We would also like to thank the organizers (IITD, Tongji, REgim, XJTU, and NTU) for allowing us to use their datasets.

REFERENCES

[1] R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," *International Journal of Compututer Vision*, vol. 128, no. 2, pp. 336–359, Feb. 2020.

[2] L.J.P. van der Maaten and G.E. Hinton, "Visualizing high-dimensional data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.