TS-Net: Combining Modality Specific and Common Features for Multimodal Patch Matching

Sovann En, Alexis Lechervy and Frédéric Jurie

Normandie Univ, UNICAEN, ENSICAEN, CNRS — UMR GREYC

firstname.lastname@unicaen.fr

Introduction

- We tackle the problem of cross-modality patch matching, *i.e.* RGB vs sketch, RGB vs near-infrared *etc.*
- In order to compare the information coming from different modalities, once has to project them onto a new subspace where the similarity can be computed, either by:
  1. using the common features among the two modalities (Siamese network)
  2. using modality-specific information (Pseudo-Siamese network)
- We show that the combination of common and modality specific features is the optimal solution (TS-Net)
- Extra supervision in the intermediate layer is used to further boost the performance
- Experimentation on three different data sets shows significant gains in performance compared to Siamese and Pseudo-Siamese approaches.
- Codes and resources available at `http://github.com/ensv/TS-Net`

Figure 1: Multi-modality patch-based matching to find corresponding RGB image from a database with the help of a partially drawn sketch.

Dataset

Table 1: Number of pairs of patches in the train, test and validation set, for each dataset (50% positive and 50% negative pairs.

Dataset	Train	Test	Validation
VeDAI	448k	128k	64k
CUHK	113k	32k	16k
NIRScene	427k	122k	61k

Performance measure: Error rate at 95% of recall.

Early or late fusion

Table 2: Performance on VeDAI data set using TS-Net. Rows: tower fusion after the feature extraction network (bottleneck layer), FC1, FC2 or FC3 of the metric layer. '1 Entropy' means there is only one classification loss at the top of the network. '3 Entropy': each sub-network also has his own classification loss. S*: Matchnet Network with the same number of parameters as TS-Net.

	3 losses	1 loss
FC3 (TS-Net)	**0.52 ± 0.07**	0.93 ± 0.05
FC2	0.62 ± 0.13	0.92 ± 0.05
FC1	0.74 ± 0.07	1.03 ± 0.06
Feature tower	n/a	1.05 ± 0.07
S*	n/a	1.01 ± 0.11

Conclusion

A novel architecture for multimodal patch matching is proposed:

- It takes advantages of both modality-specific (Siamese network) and common features (Pseudo-Siamese network)
- An additional loss helps to further boost the performance with incremental computational costs
- Experimental results demonstrate significant gains in performance compared to Siamese and Pseudo-Siamese network.

Network Architecture

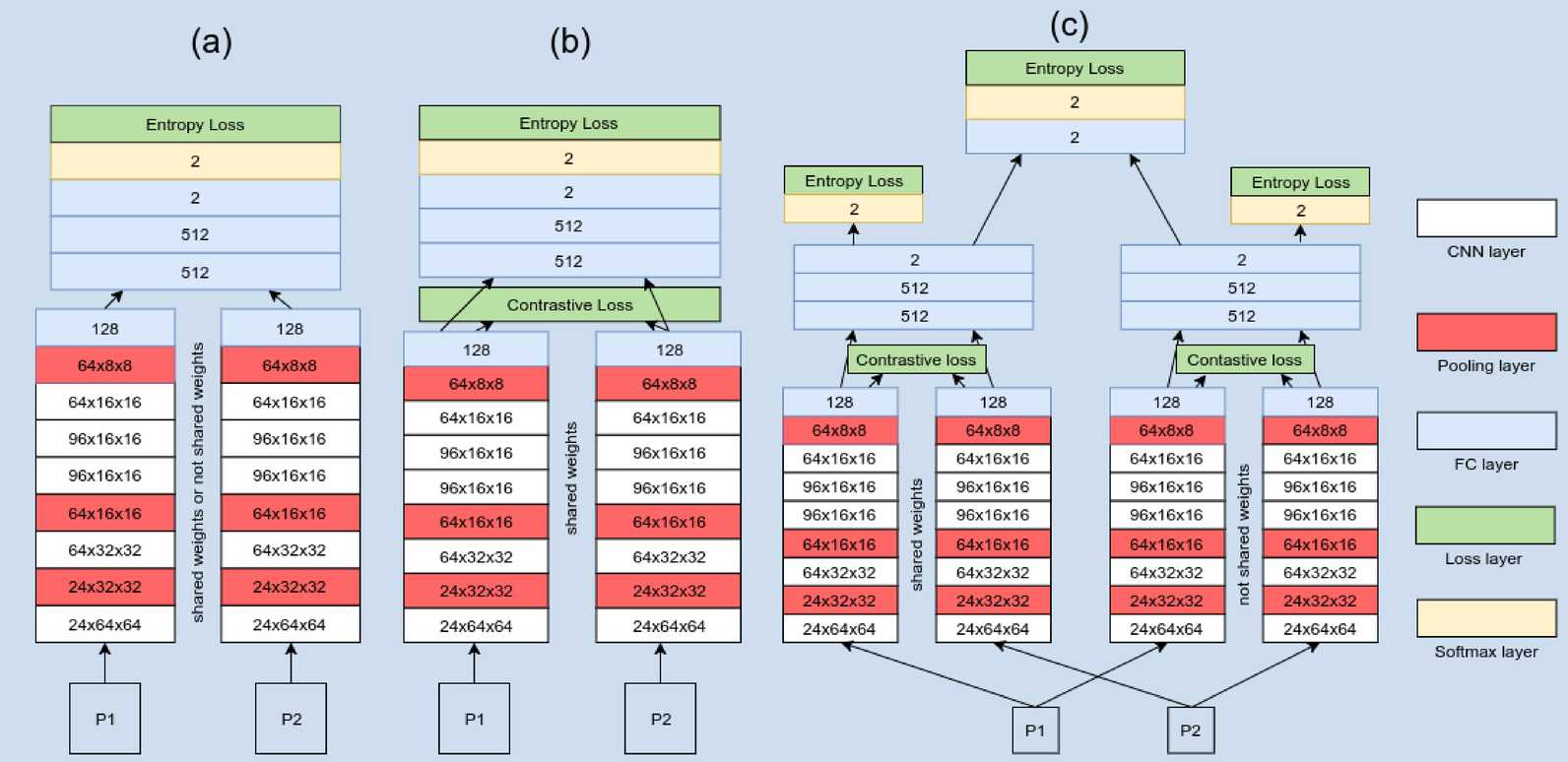

Figure 2: Illustration of our network architecture.

Experimental Results

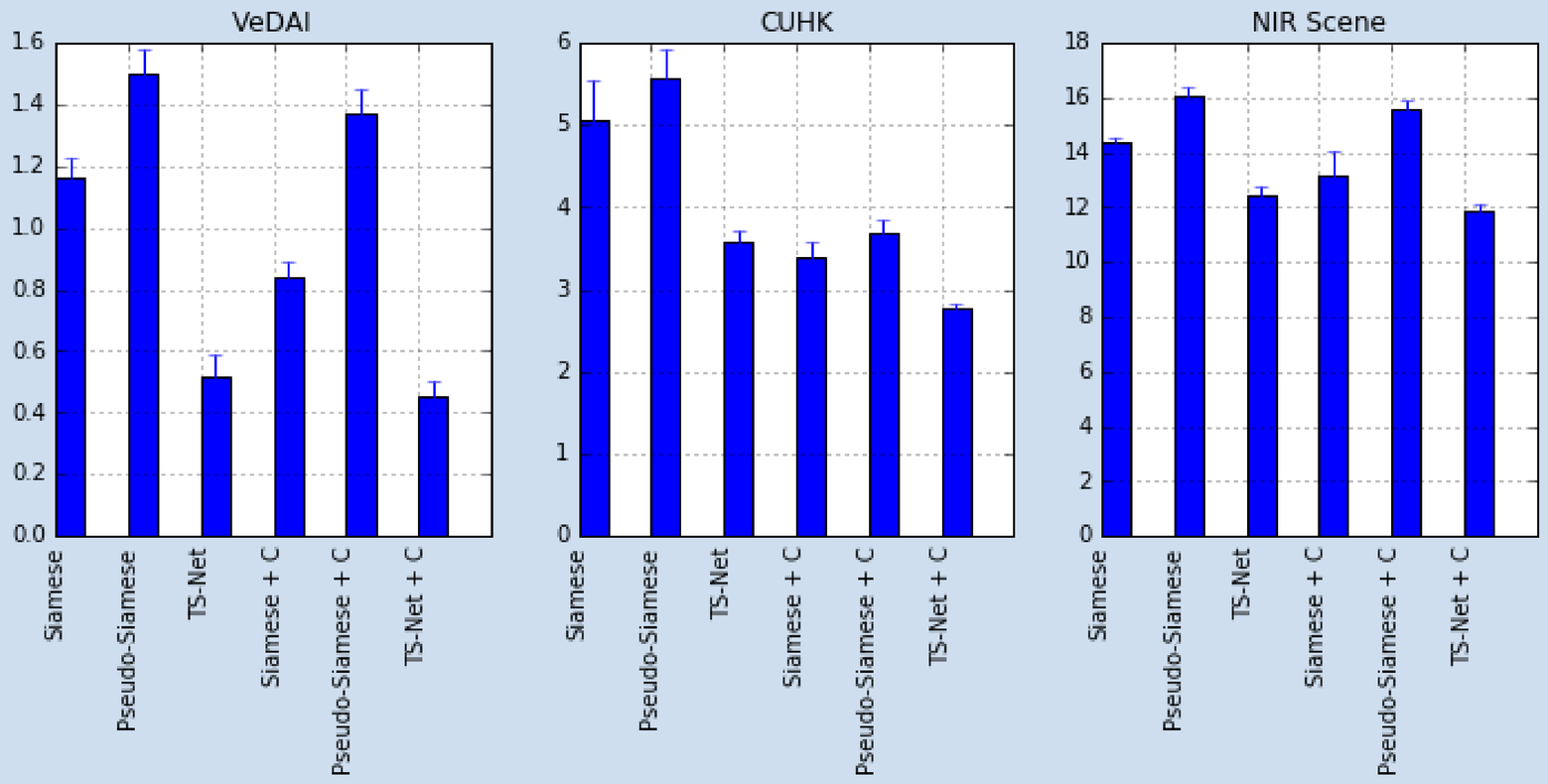

Figure 3: Performance on the 3 datasets, for Siamese network, Pseudo-Siamese network, TS-NET, without/with the additional contrastive loss (C).

Acknowledgement

This work was partly funded by the French–UK MCM ITP program and by the ANR-16-CE23-0006 program.