Automatic Feature Extraction for Social Touch Classification

Songha Ban, Ngoc NT Doan, Ruben Kole, Gökçe Kuşcu, Michon Zeegers
Department of Cognitive Science and Artificial Intelligence, Tilburg University

Motivation

Tactile Modality is vital in subtle communication and conveying emotion
→ A need for automatic social gesture recognition

Previous Studies
Classification with the Corpus of Social Touch (CoST)

- Manual feature engineering
- Highest accuracy: 60%
  → need better feature extraction to minimize possible humans' errors in hand-picked features

Goal
Automatic and inexpensive feature extraction for accurate classification

Dataset

The Corpus of Social Touch (CoST) - Jung et al. (2017)

- ❑ 14 gestures were captured.
  (Grab, hit, massage, pat, pinch, poke, press, rub, scratch, slap, squeeze, stroke, tap, tickle)
- ❑ 3 variants (gentle, normal, rough)
- ❑ The number of frames: 10 – 1747 (average: 191.78)
- ❑ Each frame consists of 64 channels (8 x 8 sensor grid)
- ❑ The channel values: 0 – 1023

Preprocessing

- ❑ Zero-padded to 1747 x 64
- ❑ Incremented by 1 before zero-padding
- ❑ Training : Validation: Test = 7 : 1 : 2

Experiment

Feature Extraction

ConvLSTM Autoencoder

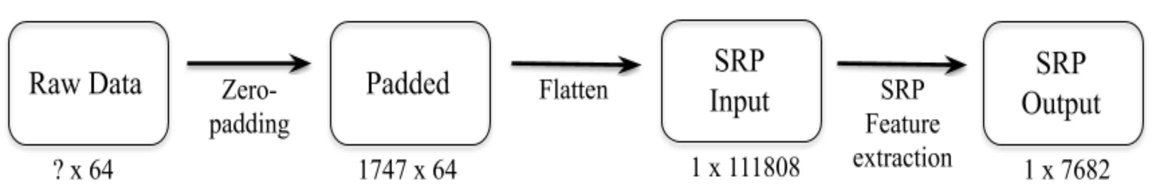
- ❑ Capture spatio-temporal correlation
- ❑ 93,640 trainable parameters
- ❑ Latent representation of size 2048

Principal Component Analysis (PCA)

- ❑ Extract the dominant patterns in the data matrix
- ❑ Output of size 6400

Sparse Random Projection (SRP)

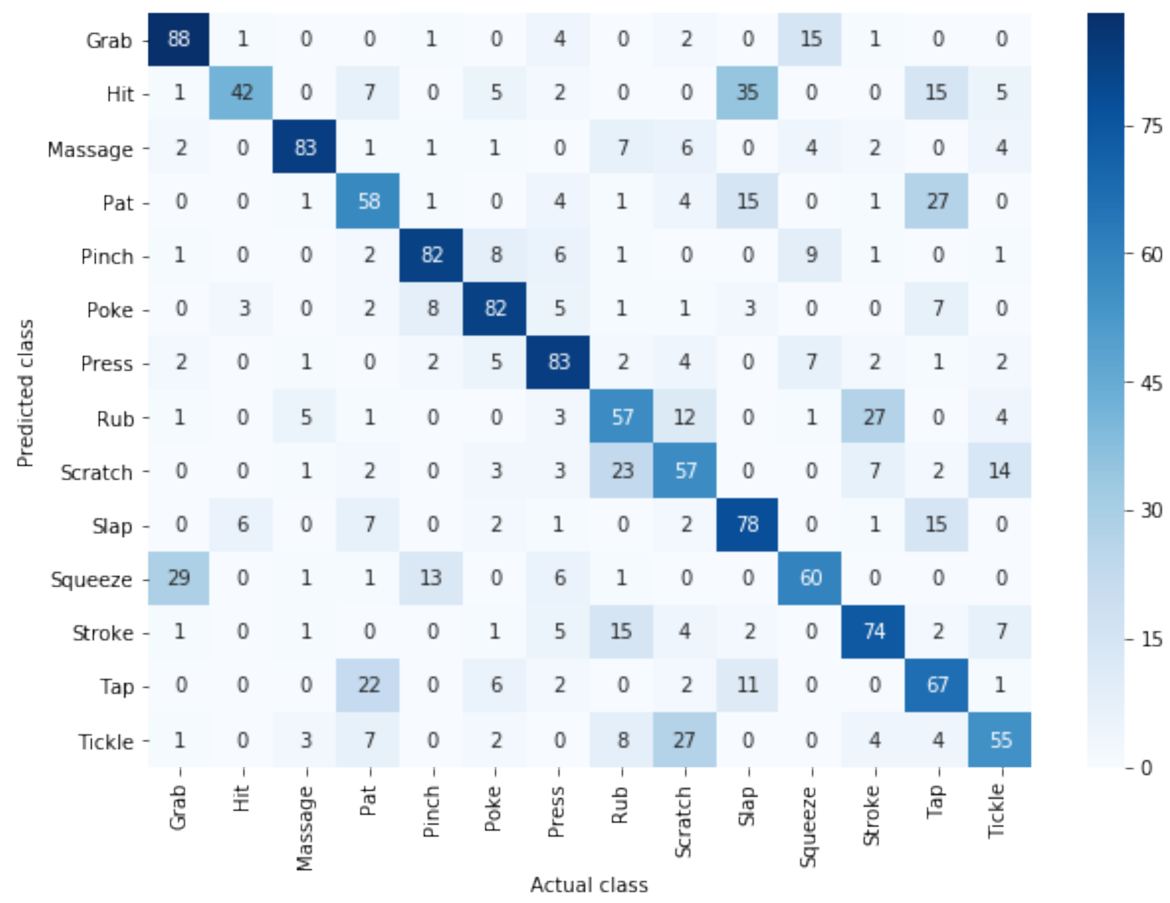
- ❑ Any high dimensional data can be converted into a lower dimensional space through random projection (Johnson-Lindenstrauss)
- ❑ Random projection matrix density: 1/sqrt(n_features)
- ❑ Output size: $n\ components \geq \frac{4\log(n\ samples)}{\varepsilon^2 - \varepsilon^3}$



Results

	ConvLSTM-AE	PCA	SRP
Output size	2048	6400	7682
Training hours	20	4	2
Training accuracy (%)	-	98.45	92.50
Test accuracy (%)	-	29.08	61.88

	All	Gentle	Normal	Rough
SRP-CNN	**.62**	**.58**	**.68**	.60
Bayesian	.57	.52	.59	.58
Decision tree	.48	.43	.49	.52
SVM linear	.59	.54	.60	**.62**
SVM RBF	.60	.54	.60	**.62**
Neural Network	.59	.52	.58	.59

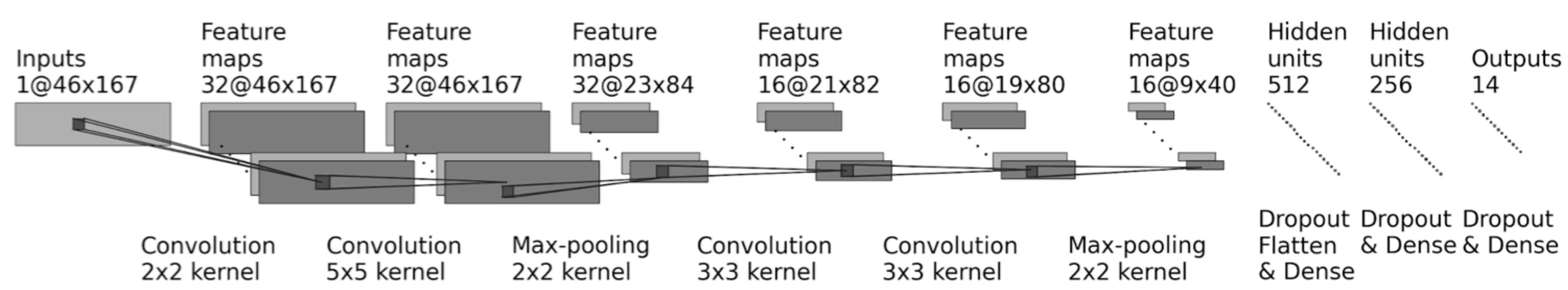


- ❑ SRP showed the best performance with test accuracy 61.88% (M=62%, SD=13%)
- ❑ Misclassification mostly due to similarity in nature between certain gestures
- ❑ Accuracy higher than the models from the previous studies
- ❑ Robust to different gesture variations

Classification

Convolutional Neural Network (CNN)

- ❑ Classified 14 different gestures (not considering variants)
- ❑ Trained 250 epochs for the best performance
- ❑ Loss function: $L(y, \hat{y}) = -\sum_{j=0}^{M}\sum_{i=0}^{N}(y_{ij} * \log(\hat{y}_{ij}))$
- ❑ Optimizer: AdaDelta



Discussion

Future Work

- ❑ Using SRP with different configurations and classifiers for a more optimal projected space
- ❑ Measuring the generalization performance of each pipeline
- ❑ Incorporating data from other modalities or other datasets of the tactile modality may enhance the classification of the gestures

Conclusion

- ❑ A more memory and computation efficient pipeline to classify social touch gestures
- ❑ A stepping stone in developing real-time touch recognition

References

[5] M. M. Jung, Mannes Poel, R. W. Poppe, and D. K. J. Heylen, "Automatic recognition of touch gestures in the corpus of social touch," *Journal on multimodal user interfaces*, vol. 11, pp. 81–96, 2017.

[8] X. SHI, Z. Chen, H. Wang, D.-Y. Yeung, W. Wong, and W. WOO, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, Eds. Curran Associates, Inc., pp. 802–810, 2015.

[9] Viorica Patraucean, Ankur Handa, and R. Cipolla, "Spatio-temporal video autoencoder with differentiable memory," 2015.

[10] Svante Wold,K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 2, pp. 37–52, 1987.

[11] Dimitris Achlioptas, "Database-friendly random projections: Johnson-Lindenstrauss with binary coins," *Journal of Computer and System Sciences*, vol. 66, pp. 671–687, 2003.

[13] P. Li, T. J. Hastie, and K. W. Church, "Very sparse random projections," pp. 287–296, 2006.

TILBURG UNIVERSITY

Understanding Society