Could you guess an interesting movie from the posters?: An evaluation of vision-based features on movie poster database

Yuta Matsuzaki, Kazushige Okayasu, Takaaki Imanari, Naomichi Kobayashi,
Yoshihiro Kanehara, Ryousuke Takasawa, Akio Nakamuraand Hirokatsu Kataoka (* equal contribution)

Department of Robotics and Mechatronics
Tokyo Denki University
{matsuzaki.y, okayasu.k}@is.fr.dendai.ac.jp
nkmr-a@cck.dendai.ac.jp

National Institute of Advanced Industrial
Science and Technology (AIST)
hirokatsu.kataoka@aist.go.jp

Problem setting

Let's try!

- Can you correct the Academy Award 2017?
- Which movie poster do you like, and why?
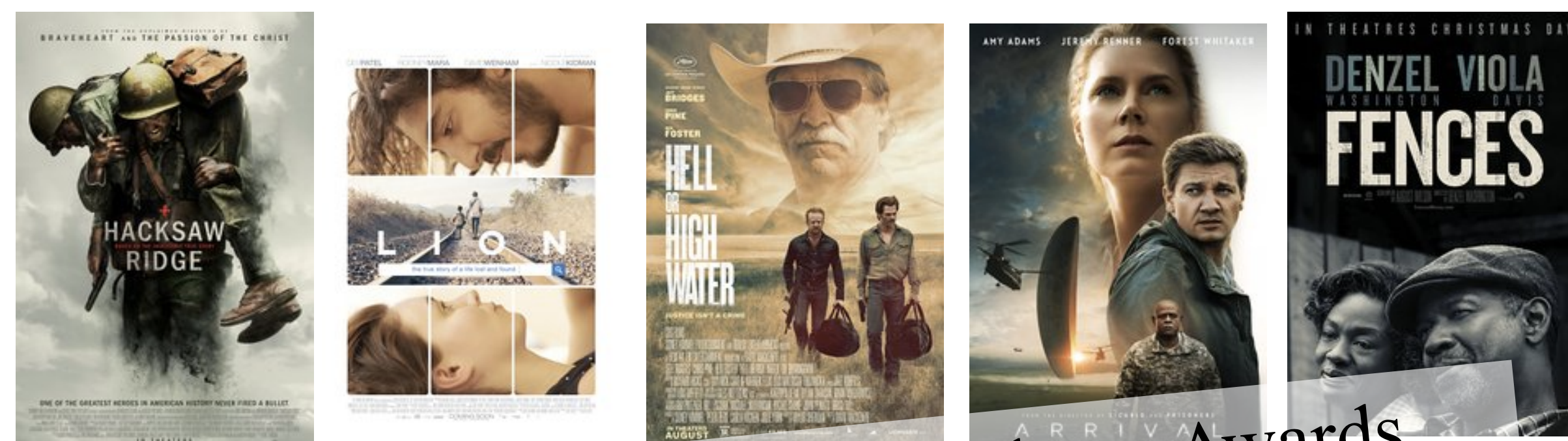
Nominated movies in 89th Academy Awards

We predict a winner in the 4 biggest film festivals from nominated movie posters

Contributions

➢ We evaluate various types of feature such as hand-craft, mid-level and deep features

➢ We have collected a database which contains,
- the nominees and winners in the 4 biggest film festivals (Academy, Berlin, Cannes, Venice)
- MPDB has 3,500+ nominate and 290+ winner works over 80 years
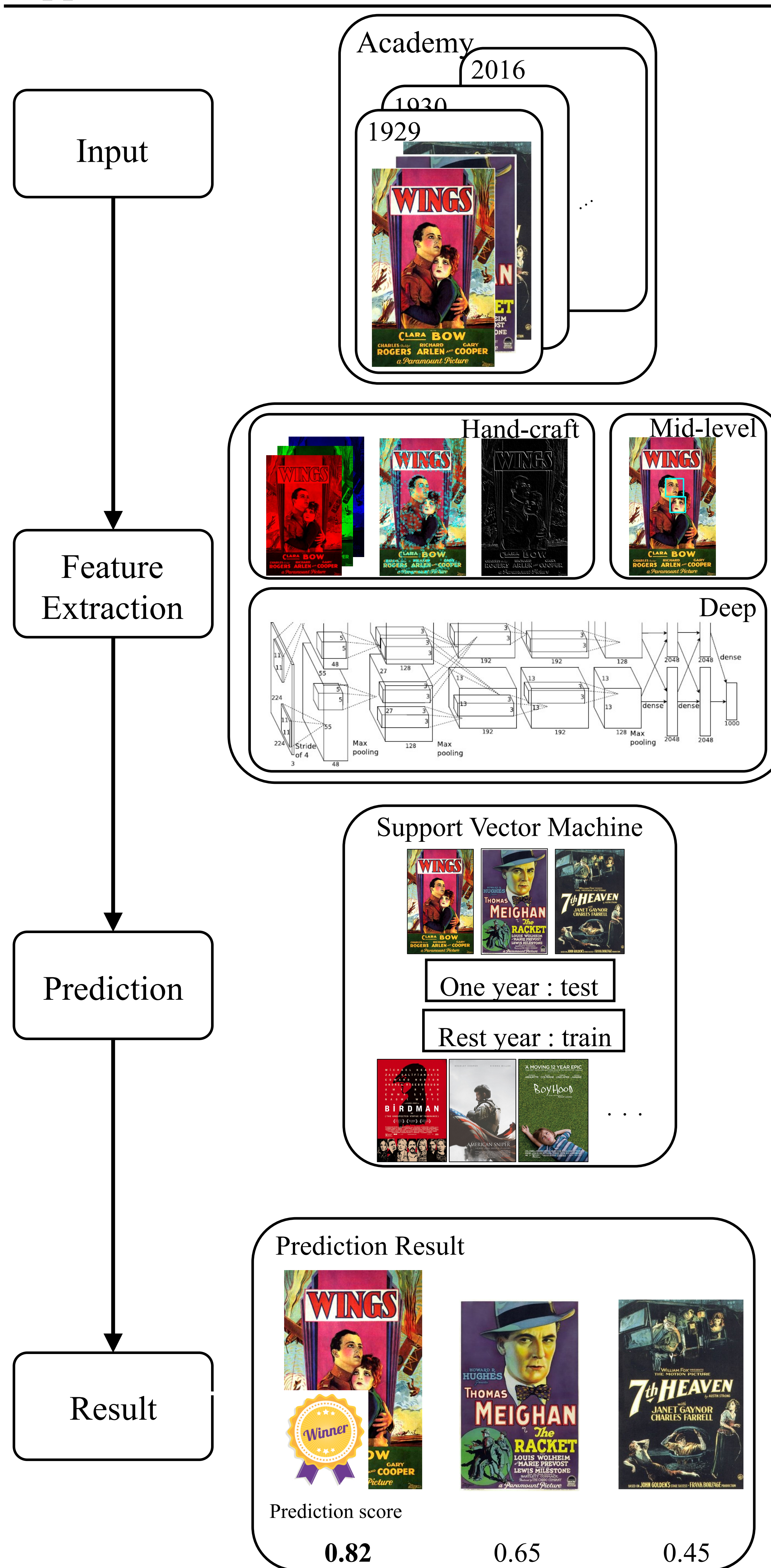
Movie Poster Database (MPDB)

Film Festival	Year	#Winners	#Nominates	Ave. #nominates at each year
Academy	1929-1932, 1934-2016	88	440	16.6
Berlin	1951-2016	63	905	6.50
Cannes	1939, 1946-1947, 1949, 1951-1968, 1969-2016	91	1335	6.38
Venice	1932, 1934-1942, 1946-1972, 1979-2016	53	869	5.75

Posters of this dataset are collected from http://www.imdb.com/

Approach

Input → Feature Extraction → Prediction → Result

Academy
2016
1930
1929
WINGS

Hand-craft   Mid-level
Deep

Support Vector Machine

One year : test
Rest year : train

Prediction Result
WINGS   MEIGHAN   7 HEAVEN
Winner
Prediction score
0.82   0.65   0.45

➢ Input
- Input certain film festival on MPDB (For the explanation, input is Academy Award as an example.)

➢ Feature extraction
We utilize various types of feature as follow
- Hand-craft feature
SIFT + BoF, HOG, CoHOG, ECoHOG, LBP, L*a*b*, GIST, Combined Handcraft feature
- Mid-level feature
PlaceNet(DeCAF), Flickr(DeCAF) Style, EmotionNet, Combined Mid-level feature
- Deep feature
AlexNet(DeCAF), VGGNet(DeCAF), Combined deep feature
(method of combination : late fusion)

➢ Prediction
We calculate prediction score by Support Vector Machine.
The parameters for identification were set to as follow.
- $C = 5.0 \times 10^{4}$
- $gamma = 1.0 \times 10^{7-5}$
- kernel = rbf
- Settings of training and testing : leave-one-year-out-cross-validation.

➢ Result
We decide a works as "winner" which have the highest prediction score at each year.
Accuracy is an average of results for all years.

Application

We predicted Oscar!
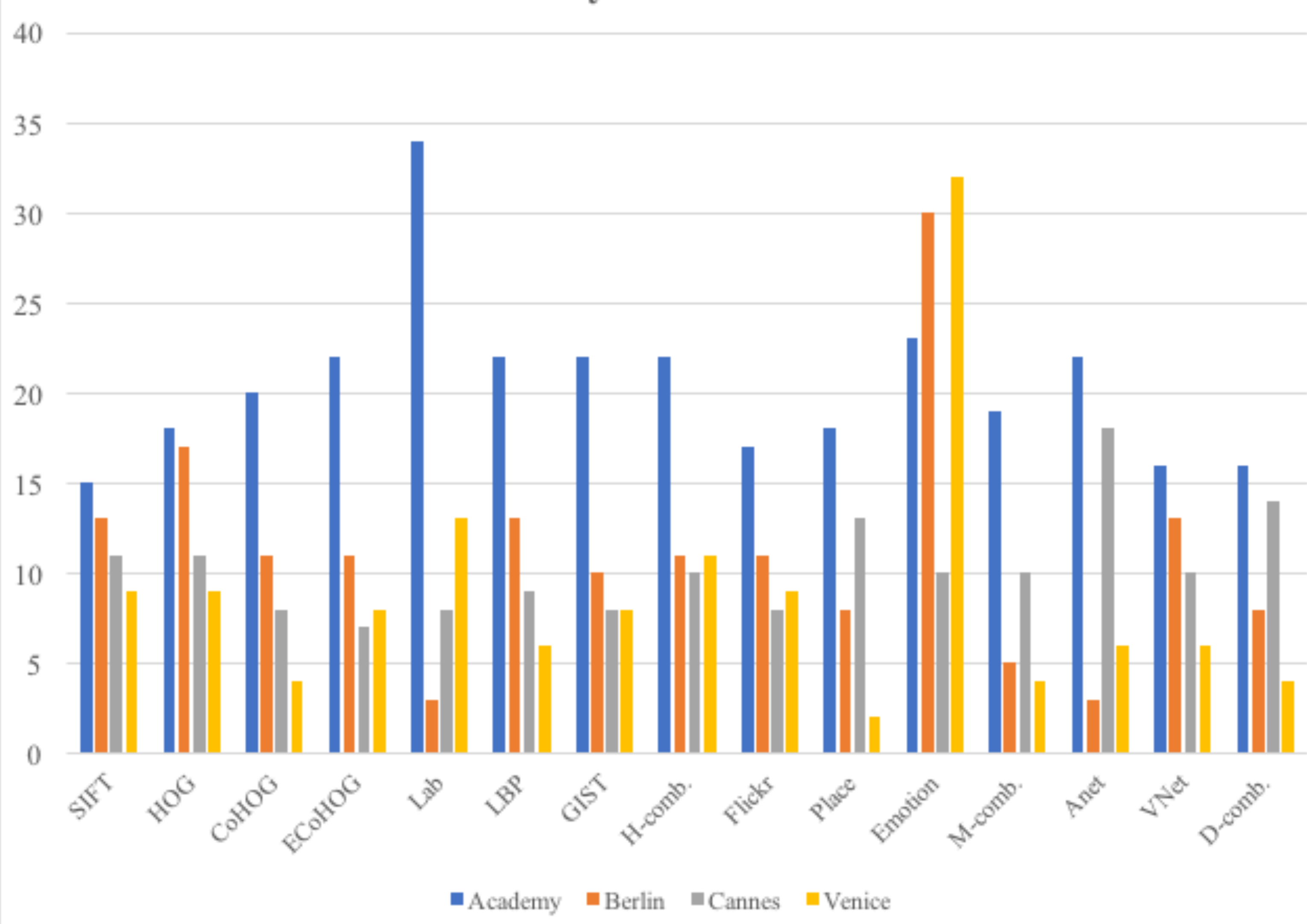
Our system got a correct answer in Academy Award 2017

Movie title	Predicted score	Rank
Moonlight [Winner]	**0.167**	**1**
Lion	0.163	2
Hell or High Water	0.162	3
Arrival	0.151	4
Hacksaw Ridge	0.142	5
Fences	0.138	6
Hidden Figures	0.114	7
Manchester by the Sea	0.112	8
La La Land	0.093	9

Result

➢ L*a*b* feature was the highest in the Academy award.
- In the Academy awards, which works selected as Winner tend to have certain colors.
  ex) red, yellow, brown, etc.,
➢ Berlin and Venice showed that the recognition rate using EmotionNet expression is the highest identification rate.
- In the Berlin and Venice, Winner works tend to draw certain facial expression at a certain position.

Accuracy in each festival

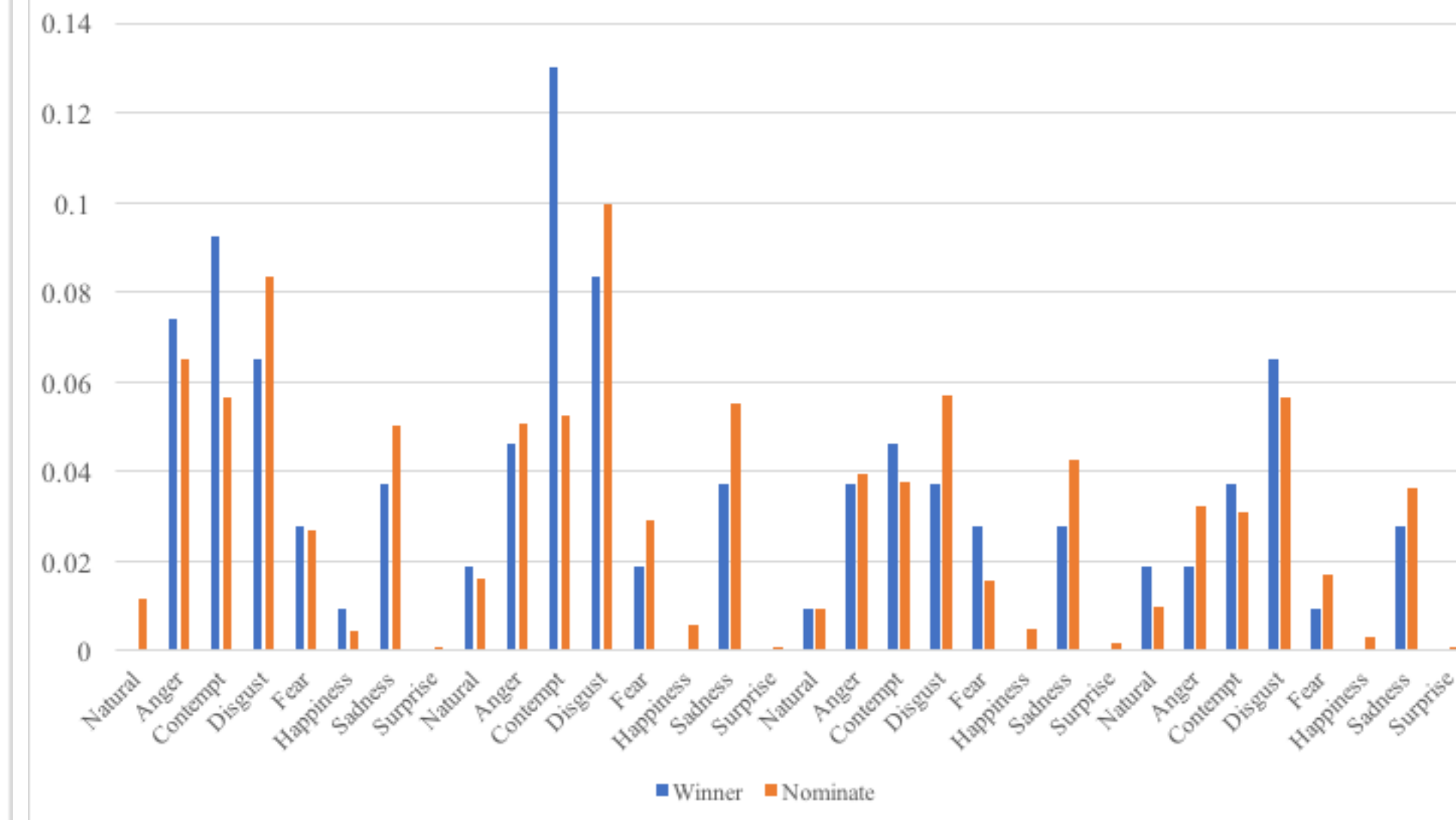Emotion histogram in Berlin

Emotion histogram in Venice

References

[1] T. Watanabe, et al.: "Co-occurrence Histograms of Oriented Gradients for Pedestrian Detection," Proceedings of the 3rd Pacific Rim Symposium on Advances in Image and Video Technology, pp. 37--47, 2009.

[2] H. Kataoka, et al.: "Extended Feature Descriptor and Vehicle Motion Model with Tracking-by-detection for Pedestrian Active Safety," IEICE Transactions on Information and Systems, Vol.E97-D, No.2, 2014.

[3] F. Palermo, et al.: "Dating historical color images," In Proceedings of the European Conference on Computer Vision, pp.499--512, 2012.

[4] T. Ojala, et al.: "A Comparative Study of Texture Measures with Classification Based on Feature Distributions," Pattern Recognition, vol.29, pp.51--59, 1996.

[5] A. Oliva, et al.: "Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope," International Journal in Computer Vision, vol.13, no.42, pp.143--175, 2001.

[6] S. Karayev, et al.: "Recognizing Image Style," Proceedings of the British Machine Vision Conference, 2013.

[7] B. Zhou, et al.: "Learning Deep Features for Scene Recognition using Places Database," Advances in Neural Information Processing Systems, 2014.

[8] B. Kennedy, A. Balint, 2016. Github: {https://github.com/co60ca/EmotionNet}

[9] A. Krizhevsky, et al.: "ImageNet Classification with Deep Convolutional Neural Networks," Advances in Neural Information Processing Systems, 2012.

[10] K. Simonyan, et al.: "Very Deep Convolutional Networks for Large-Scale Image Recognition," International Conference on Learning Representations}, 2015.