MediaEval 2016: A multimodal system for the Verifying Multimedia Use task

Cédric Maigrot, Vincent Claveau, Ewa Kijak, Ronan Sicre

MᴇᴅɪᴀEᴠᴀʟ 2016: A ᴍᴜʟᴛɪᴍᴏᴅᴀʟ sʏsᴛᴇᴍ ғᴏʀ ᴛʜᴇ Vᴇʀɪғʏɪɴɢ Mᴜʟᴛɪᴍᴇᴅɪᴀ Usᴇ ᴛᴀsᴋ

Cédric Maigrot   Vincent Claveau   Ewa Kijak   Ronan Sicre

{firstname}.{lastname}@irisa.fr

Why use a multimodal system ? Because there are several types of hoax !

➡ False information present in the text content ➡ Forged image ➡ Image reused for an other event

Global Hypotheses

➡ Prediction is first made at the image-level, then propagated to the tweets that contain the image
➡ Translation if the detected language is different than english

Text-based approach (run-T)

Detect if the message is style-wise similar to known hoax

➡ Capture similar comments between an unknown image and an image from the training set (e.g. *It's photoshopped*) and similar genres of comments (e.g. presence of smileys)
➡ Prediction made by a k-Nearest-Neighbor approach (in this case $k = 1$)

Source-based approach (run-S)

Detect if the message is related to a trustworthy source

➡ 2 type of sources searched: news-related organisms (e.g. press agencies) and explicit citations of the source of the image (e.g. the pattern *photographed by + Name*)
➡ Predict *real* if a trustworthy source is detected, *fake* else
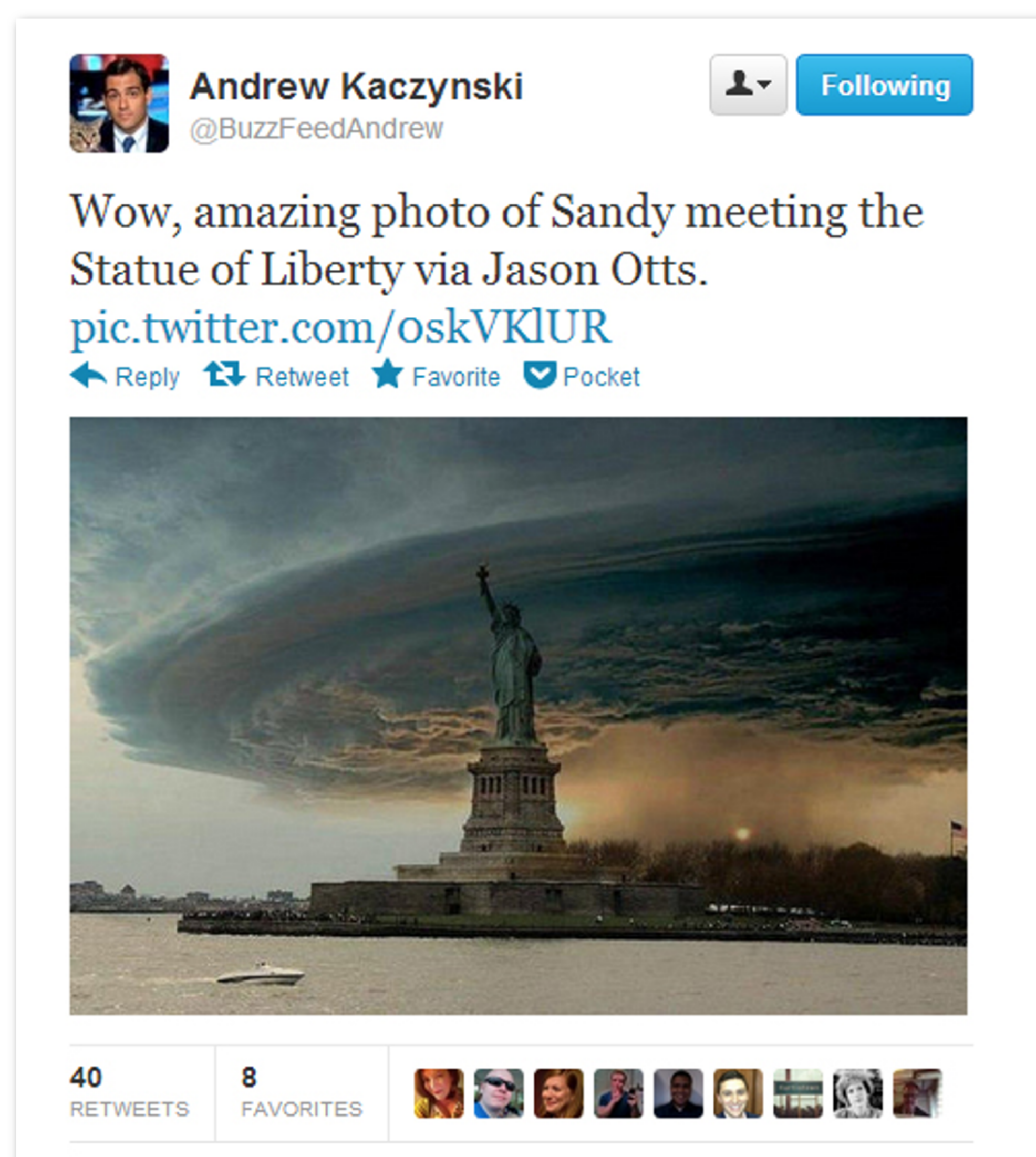
Example

Andrew Kaczynski
@BuzzFeedAndrew
Following

Wow, amazing photo of Sandy meeting the Statue of Liberty via Jason Otts.
pic.twitter.com/oskVKlUR

Reply Retweet Favorite Pocket

40 RETWEETS   8 FAVORITES

Image-based approach (run-I)

Detect a known image

➡ Compare an unknown image to an image database of 8 000 known images (7 500 fake and 500 real images)
➡ Database images extracted from 5 specialized websites
➡ Description of an image by a deep CNN layer output (4096-dimensional descriptor)
➡ Predict *real* (resp. *fake*) if a *real* (resp. *fake*) similar image is found in the database, *uncertain* else
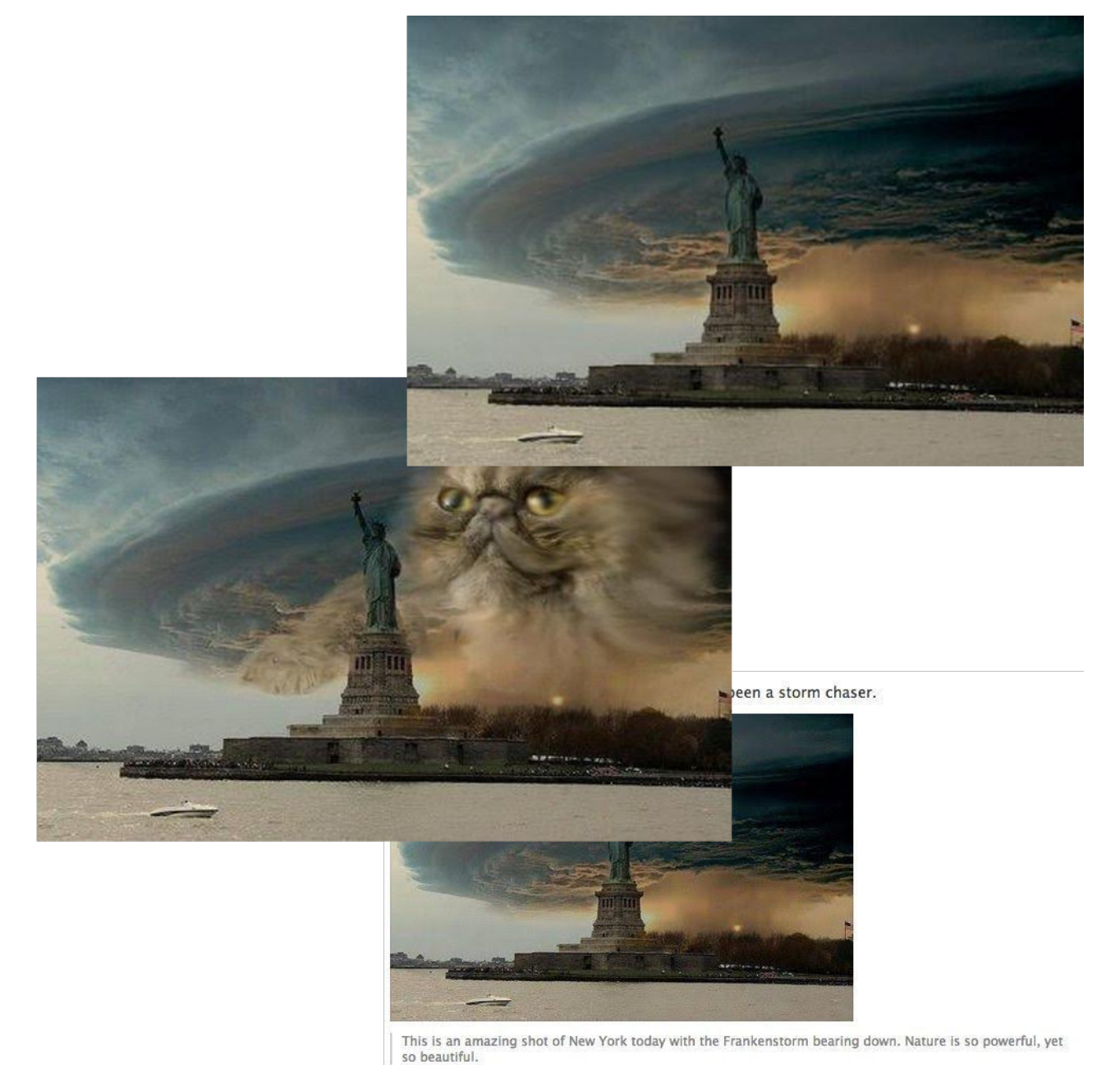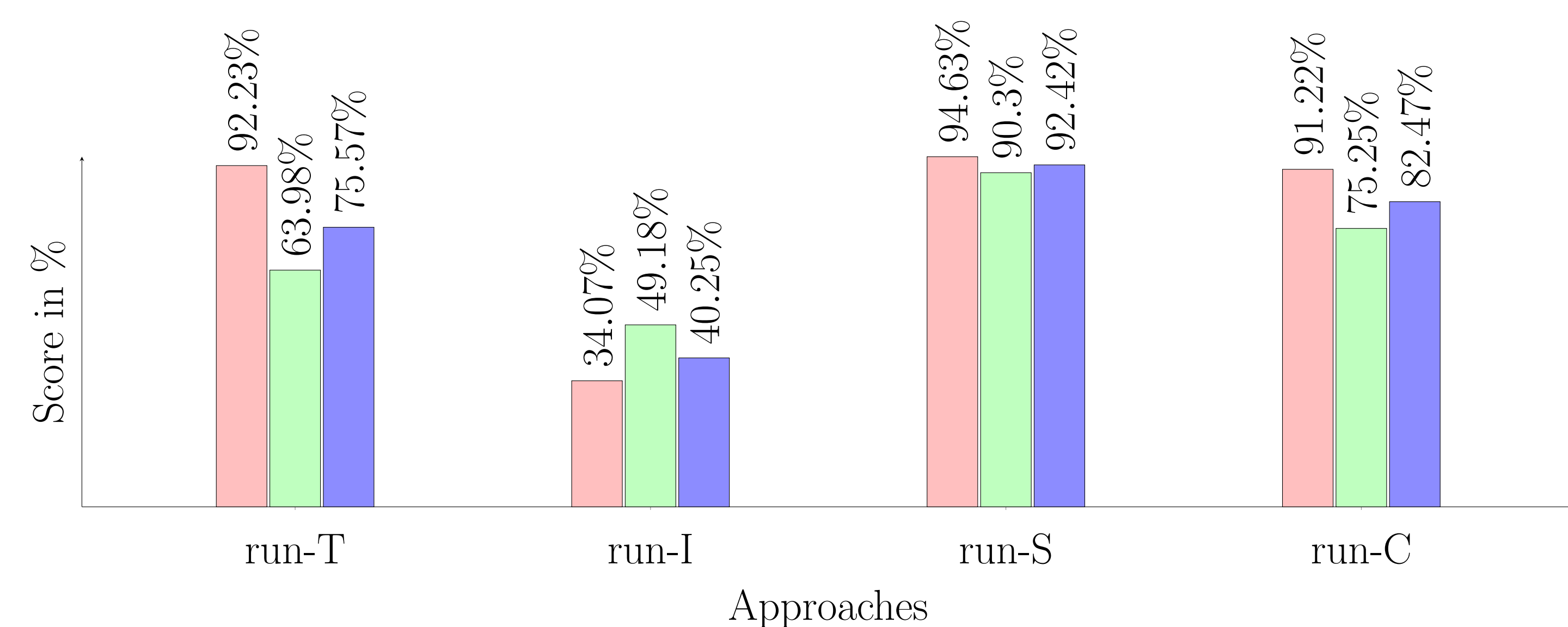
Combination approach (run-C)

Combine the three previous predictions

➡ Late fusion: learn the best combination
➡ Boosting algorithm (adaboost.MH, parameters of the machine learning algorithm are set by cross-validation on the training data)

Results

Score in % / Approaches

	run-T	run-I	run-S	run-C
	92.23%	34.07%	94.63%	91.22%
	63.98%	49.18%	90.3%	75.25%
	75.57%	40.25%	92.42%	82.47%

➡ 2 228 messages to classify, corresponding to 130 images
➡ 86 % to the test tweets are associated with one or more images (the rest is associated with video)

Conclusion

➡ Text-based approach: competes with the source-based approach in terms of recall but tends to classify every tweet as *fake*
➡ Image-based approach: low precision compared with estimations on the training set. This may be due to: (1) small and unbalanced reference database; (2) original image and forged ones are sometimes very similar; (3) presence of stamps
➡ Combination-based approach: does not offer any gain due to overfitting

Acknowledgements