Universität Augsburg
Fakultät für Angewandte Informatik

Towards Demystifying Subliminal Persuasiveness: Using XAI-Techniques to Highlight Persuasive Markers of Public Speeches

Klaus Weber, Lukas Tinnes, Tobias Huber, Alexander Heimerl, Eva Pohlen, Marc-Leon Reinecker, Elisabeth André

AAMAS 2020 / EXTRAAMAS 2020 - Auckland New Zealand - 09 Mai 2020 - 13 Mai 2020

Background 1/2

- In public speeches: People without previous knowledge can be persuaded very easily.
- However, not only logical arguments are used
  - Non-verbal play an important role.
- Presenting content-wise identical arguments in different ways can have a different effect on the audience's opinion.
- People are not aware of this subliminal persuasion process.

Background 2/2

- Understanding these cues bears several advantages:
  - People can learn to behave differently -> more persuasive.
  - People can use this understanding for the development of persuasive robots and agents.

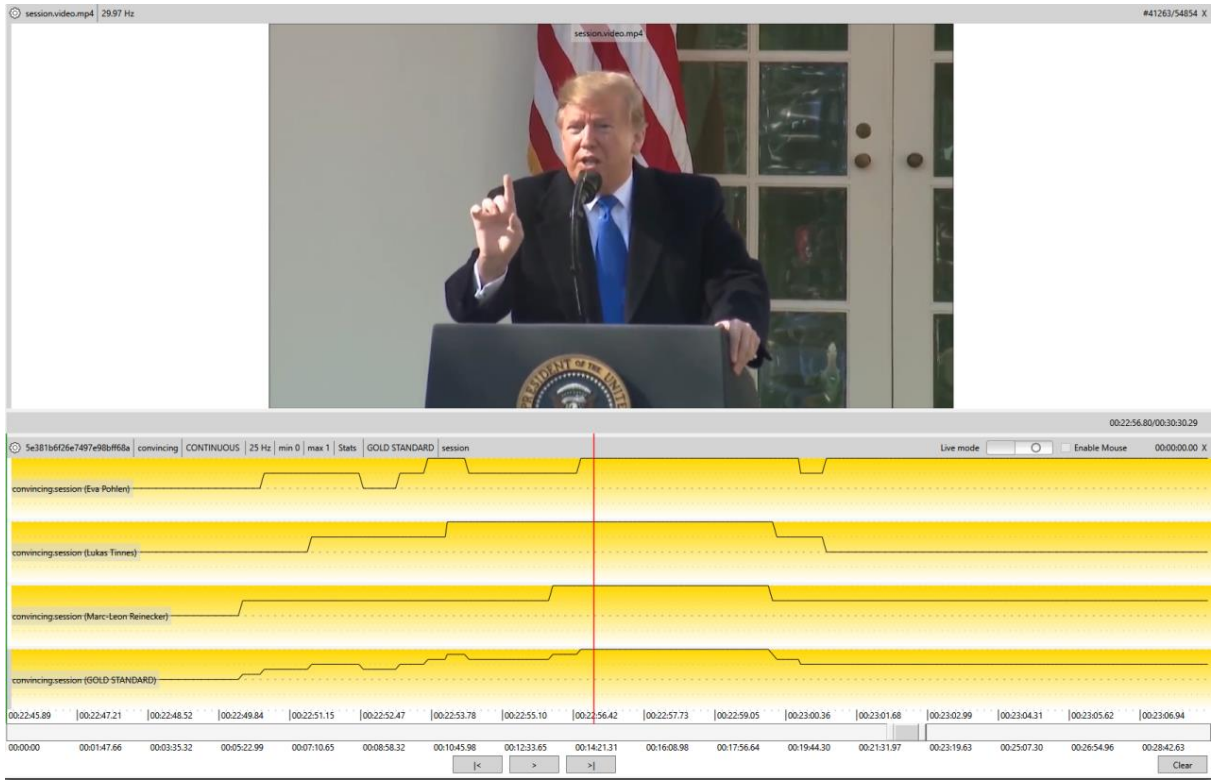
- **Overall Goal:** Raising awareness of subliminal persuasion using XAI techniques!

Approach

1. Annotating a video of a public speech based on video and audio data.
2. Training of a neural network based in video data only.
3. Post-Hoc Analysis of the trained network using XAI.

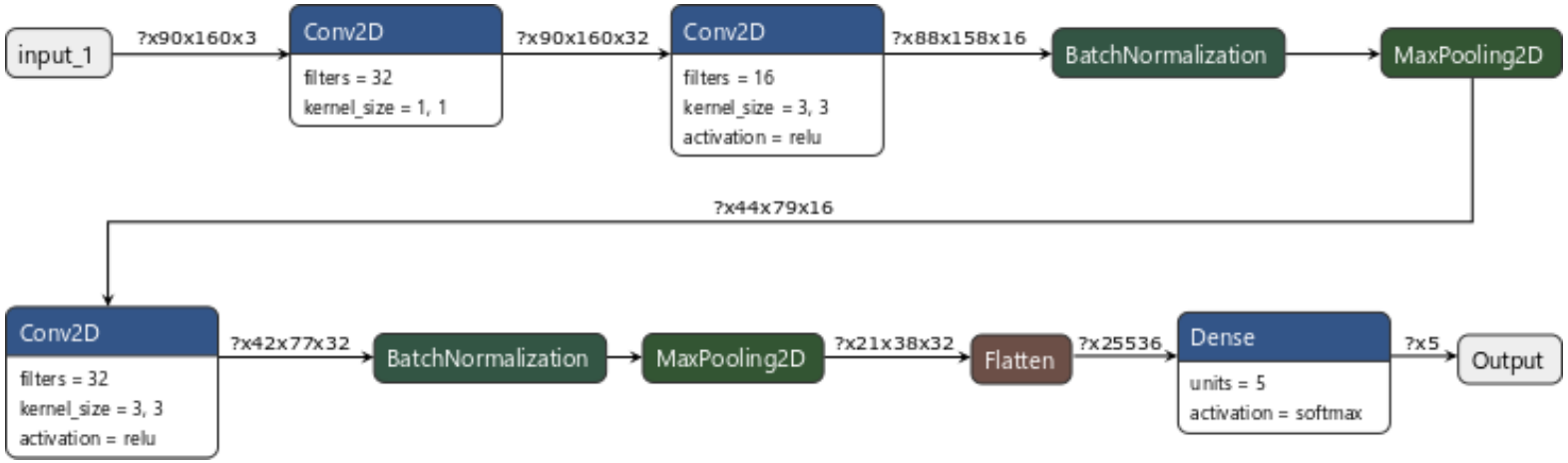
Concept 1/2

- Annotated video of a political speech with respect to the perceived persuasiveness:
  - Very convincing
  - Moderately convincing
  - Neutral
  - Hardly convincing
  - Not convincing
- Three experienced labelers:
  - Inter-rater agreement: 0.77 (Cronbach‘s-Alpha)



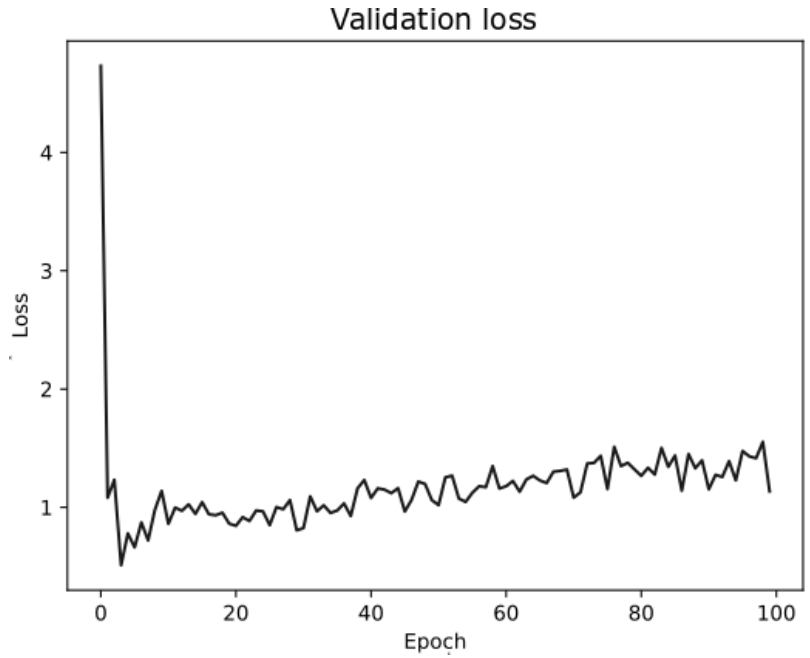
Concept 2/2

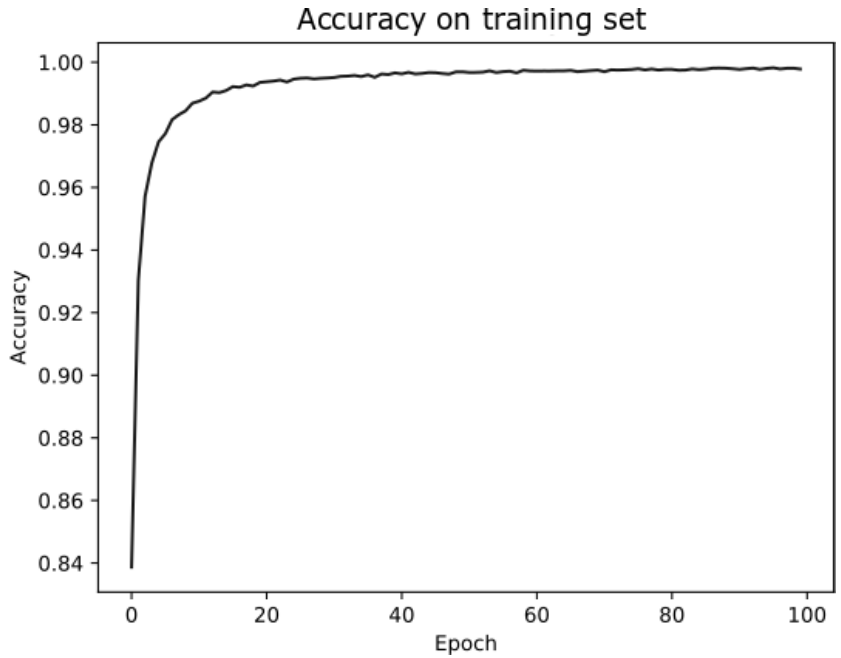
- Training of convolutional neural network using the extracted video frames only and the annotated data to predict the perceived persuasiveness.
  - ~ 50,000 frames
  - Sample rate: 25Hz
  - Frames downsampled to 190x60
  - batch size: 32
  - Optimizer: Adamax
  - Batch normalization to tackle overfitting



Training performance 1/3

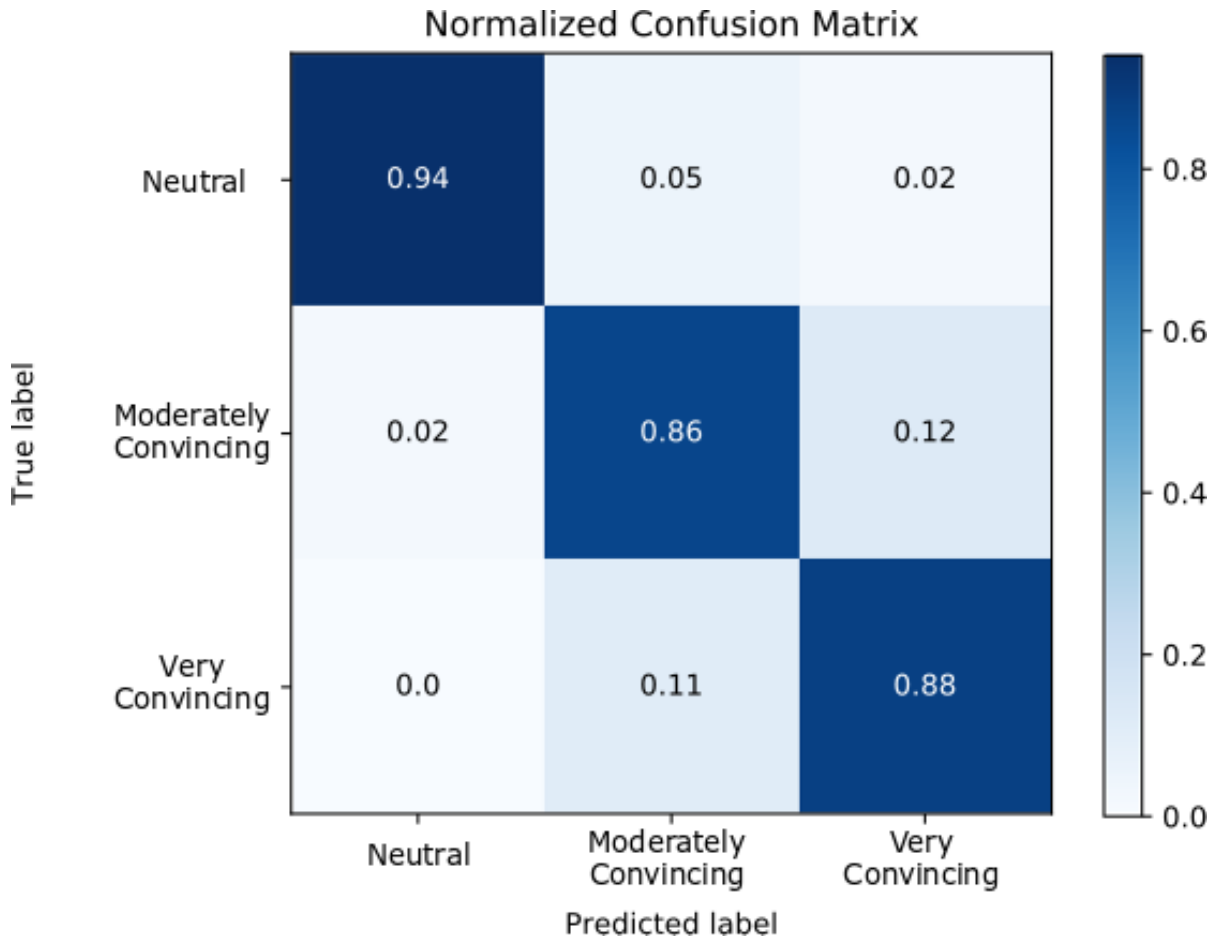
- 100 episodes:
  - Slight overfitting.
  - Analyzed the model after 20 episodes of training.





Training performance 2/3

- Confusion matrix computed on the training data to ensure that the network is sufficiently correct on the learned samples.
  - Video only consisted of only three classes.
  - High accuracy over existing classes.



Training performance 3/3

Measure	Class Neutral	Moderately Convincing	Very Convincing
Precision	0.93	0.93	0.77
Recall	0.94	0.86	0.88
F1-Score	0.93	0.89	0.82

Table 1. Precision, Recall and F1-Score for all existing classes.

Post-Training-Analysis: What does the network look at?

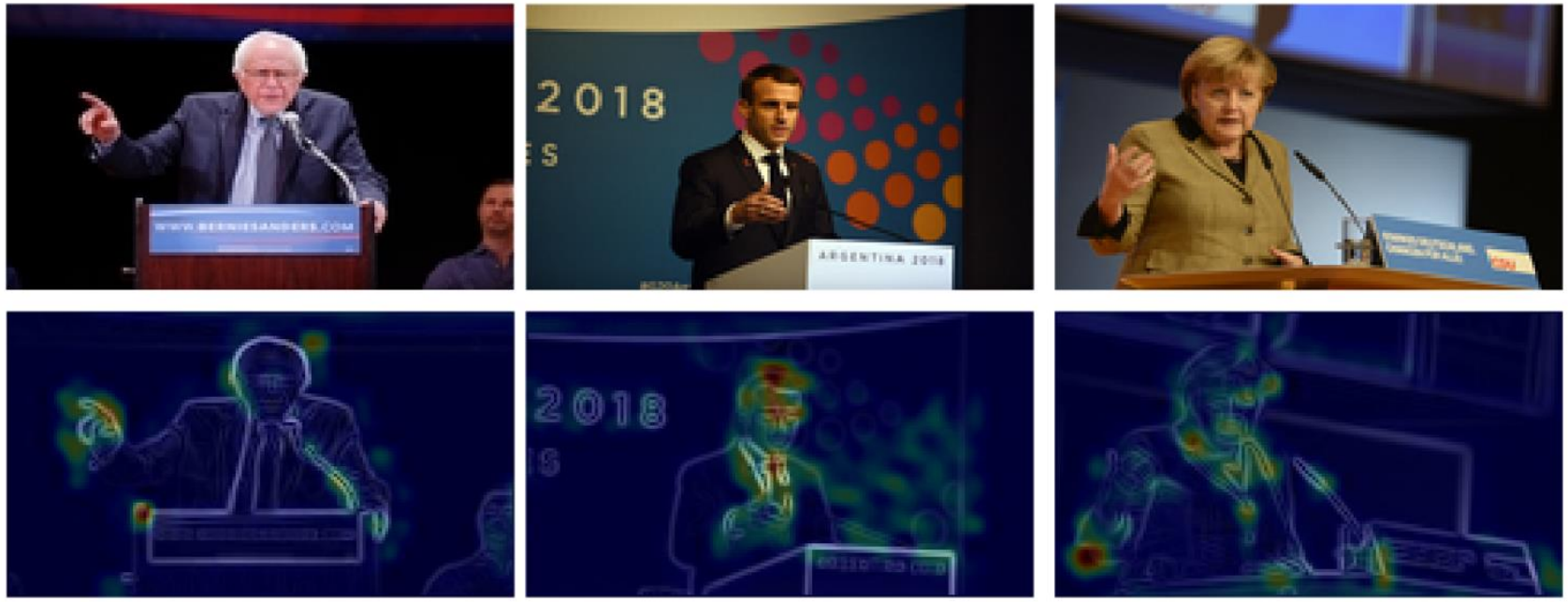
- Network only sees the images without audio signal.
  - Does the network look at cues that we generally consider as persuasive?
  - Are those cues in line with existing literature.
- Applying of two XAI techniques:
  - Grad CAM
  - LRP
- We chose these two to…
  - ...get explanations at the end of the model (grad-CAM).
  - ...get explanations at the beginning of the model (LRP).

Post-Training-Analysis: grad-CAM

- Visualizations of three classes (FLTR): ***neutral***, ***moderately convincing***, ***very convincing***.
  - Network focuses on person's contours.
  - Neutral class: No persuasive indicators.
  - Networks follows hand and arms.

Post-Training-Analysis: grad-CAM - Different speakers

- Network tested on different speakers (FLTR): ***Bernie Sanders***, ***Emanuel Macron***, ***Angela Merkel***.
  - Network still focuses on person's face and hands.
  - Picture of Macron: Reveals that the network seems to focus on skin-related areas.



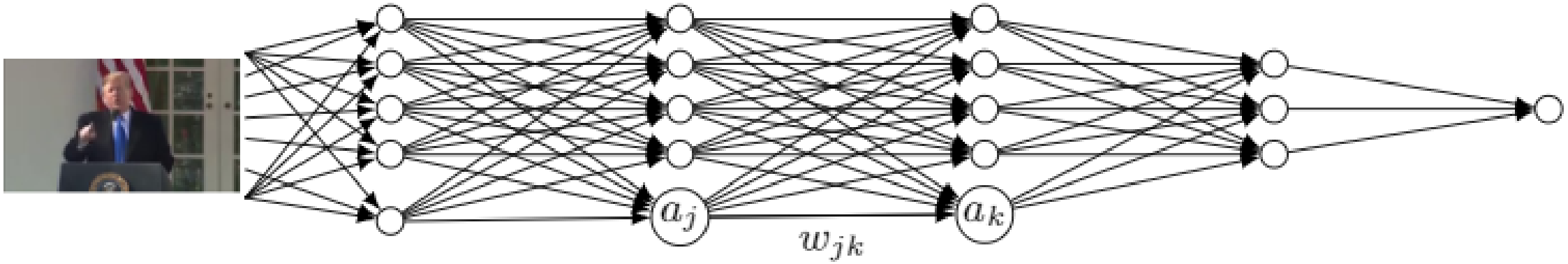
Post-Training-Analysis: LRP 1/2

- LRP-Method: ***z⁺-rule***.
  - Assigns relevance Value to each neuron k in the network:

$$R_k = \begin{cases} a_k & \text{if } k = argmax\{a_k\} \\ 0 & \text{if not.} \end{cases}$$

  - Relevance gets propaged back:

$$R_j = \sum_k \frac{(a_j w_{jk})^+}{\sum_j (a_j w_{jk})^+} R_k$$

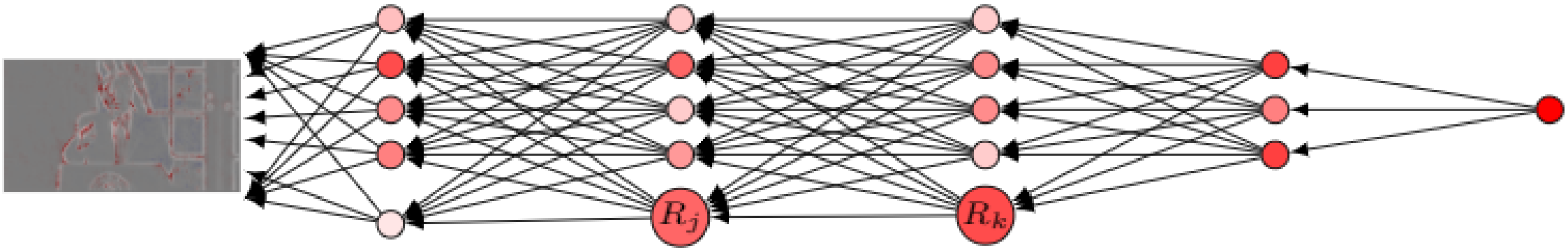


Post-Training-Analysis: LRP 1/2

- LRP-Method: ***z⁺-rule***.
  - Assigns relevance Value to each neuron k in the network: .
  - Relevance gets propaged back:

$$R_k = \begin{cases} a_k & \text{if } k = argmax\{a_k\} \\ 0 & \text{if not.} \end{cases}$$

$$R_j = \sum_k \frac{(a_j w_{jk})^+}{\sum_j (a_j w_{jk})^+} R_k$$



Post-Training-Analysis: LRP 2/2

- Visualizations of three classes (FLTR): ***neutral***, ***moderately convincing***, ***very convincing***.
  - Network focuses on person's contours.
  - Networks follows hand and arms.

Limitations 1/2

- **Limited Training corpus**
  - Only consisted of 50,000 samples of the same person.
  - In this regard, Learning results should be interpreted with some care.
  - It should not be considered to be a general predictor for persuasiveness.
- **Annotated Data and Annotation Process**
  - Only consisted of three classes.
    - neutral
    - moderately convincing
    - very convincing
  - Network has not learned what not convincing looks like.
  - Persuasion subject to person’s own opinion.
    - If annotators annotated the perceived persuasiveness or the intensity of the body language requires further analysis.

Limitations 2/2

- **Model**
  - Current approach uses a CNN.
  - Prediction made on single image only.
  - There are persuasive indicators that depend on movement:
    - Speed of hand gestures.

Summary and Conclusion

- Explored an approach to highlight persuasive markers of public speeches.
  - Evidence that non-verbal cues are very important when persuading people.
  - People are often not aware of subliminal persuasion.
- Trained a CNN to predict the perceived persuasivness.
  - Based on image input only.
- Applied two XAI-techniques: Grad-CAM + LRP …
  - … showing that the network focues on the person’s arms, hands and contours.

Future Work

- Extending corpus with other speakers …
  - ... to obtain more generalized training results.
  - … to get more detailed visualization of different classes.
- Looking for existing persuasion corpora.
- Using standard network architectures, such as VGG.
- Applying different other XAI techniques.