Exploiting the Logits: Joint Sign Language Recognition and Spell-Correction

Christina Runkel
Stefan Dorenkamp
Hartmut Bauermeister
Michael Moeller

UNIVERSITÄT SIEGEN

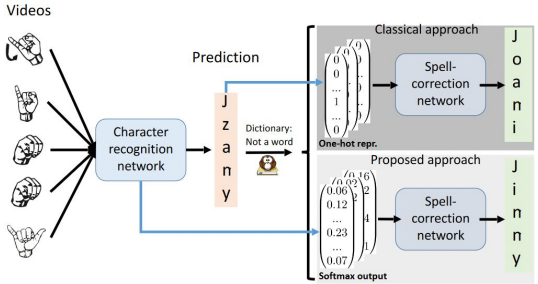
Sign Language Recognition

The automatic recognition and translation of sign language with cameras on mobile devices bares a great potential to impact our social life. While deep learning techniques have revolutionized the field of computer vision, significant challenges remain in the automatic analysis of video data due to their high dimensionality as well as the comparably scarce training data available to train activity recognition systems. We consider the recognition of letters of the German finger alphabet. Since this also includes moving gestures, we analyze videos instead of images or segregated frames.

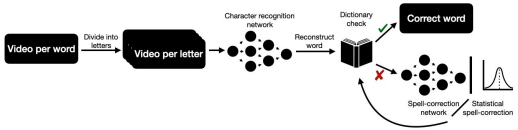
Overview of German finger alphabet [1]

Proposed Approach

Due to scarce training data, we propose a two fold approach using a character recognition network followed by a spell-correction network.



As the classification network achieves a top 5 accuracy of 98%, instead of one-hot representations, the softmax outputs of the first network are used as inputs for the spell-correction network such that the overall workflow is as follows:



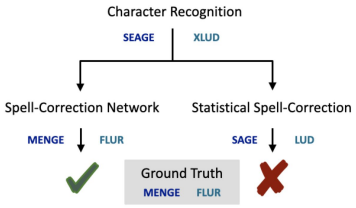
Quantitative Evaluation

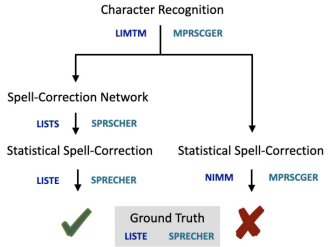
The evaluation was performed on different variants of spell-correction. The following tables show the results for hardmax and softmax inputs. Furthermore, we distinguish between statistical spell-correction methods, the spell-correction network and our approach which uses the spell-correction network as well as an additional statistical spell-correction. With regard to word accuracy, it is also important to note that some words despite not being corrected completely, are often readable by humans.

	Inputs	Character Accuracy	Word Accuracy
Spell-correction network	**Hardmax**	0.81	0.36
	Softmax	**0.85**	**0.44**
Spell-correction network + statistical spell-correction	**Hardmax**	0.81	0.63
	Softmax	**0.89**	**0.75**

	Character Accuracy	Word Accuracy
Statistical spell-correction	0.77	0.61
Spell-correction network	0.85	0.44
Spell-correction network + statistical spell-correction	**0.89**	**0.75**

Qualitative Evaluation





Conclusion

Exploiting the softmax probabilities of a character classifier can yield systematic improvements in the spell-correction. Attention has to be paid to the problem of overfitting the spell-corrector to specific examples of softmax probabilities from the classifier, such that mostly training on hardmax outputs but exploiting softmax outputs during inference seems to be favorable - at least in our case of rather scarce training data. By combining learning and statistical methods our final pipeline achieves word accuracies that are up to 25% higher than purely using a statistical approach.

References

[1] *Deutsche Gebärdensprache (DGS) und Fingeralphabet*. https://www.aktion-mensch.de/dafuer-stehen-wir/was-ist-inklusion/deutsche- gebaerdensprache.html. – Accessed: 20.09.2019