

Is multitask learning always better?

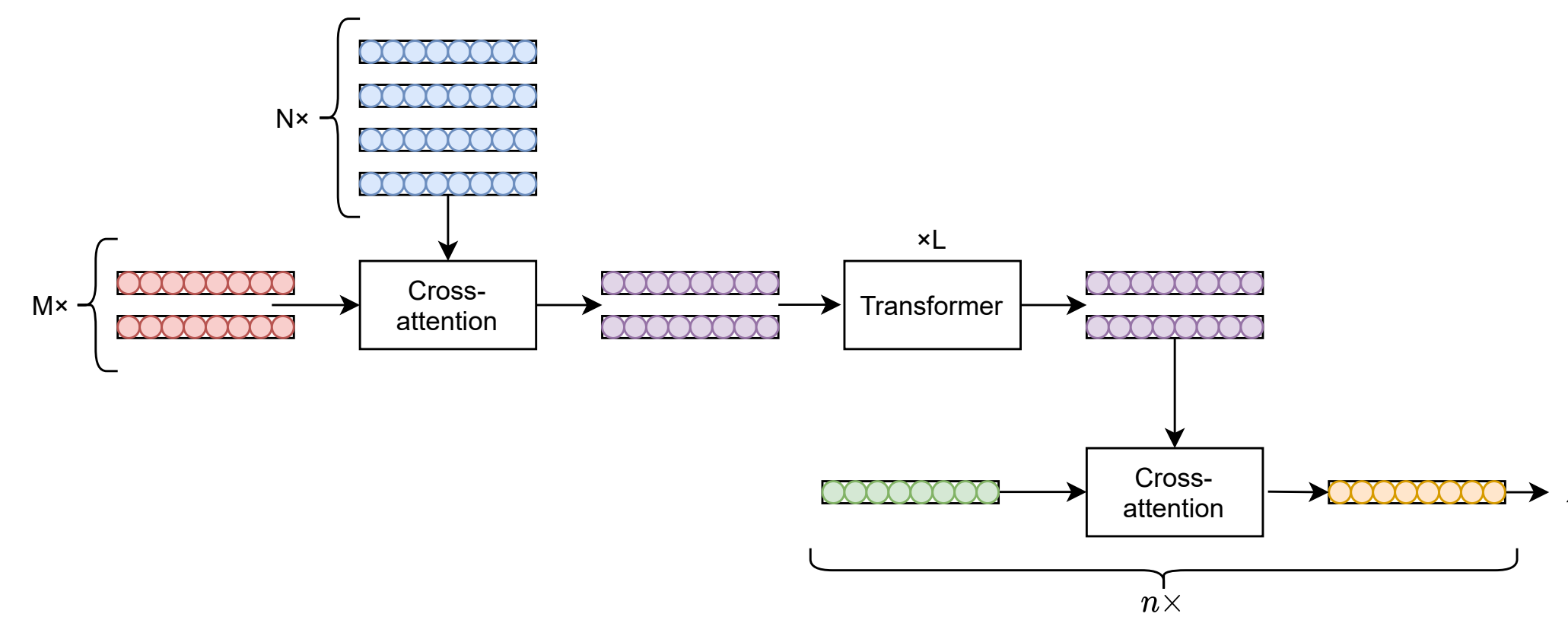
Alexander Mattick, Martin Mayr, Andreas Maier, and Vincent Christlein

Pattern Recognition Lab, Friedrich-Alexander University Erlangen-Nürnberg, Erlangen, Germany

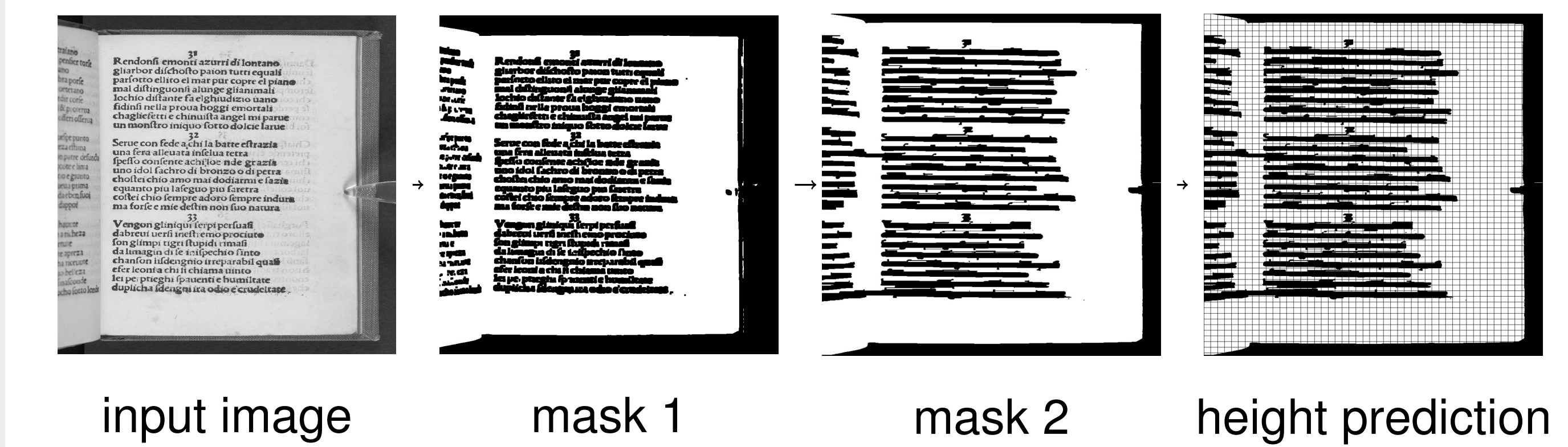
Testing Methodology

- Obtain tasks “font”, “script”, “location”, “date” from [1]
- Normalize documents based on fontsize
- Train ResNet[2] and Perceiver[3] models on
 - ▶ All 4 tasks
 - ▶ Individual tasks
 - ▶ Task pairs
- Analyze class distributions
- Analyze individual task accuracies

Perceiver Architecture



Task-independent Normalization



ResNet Results

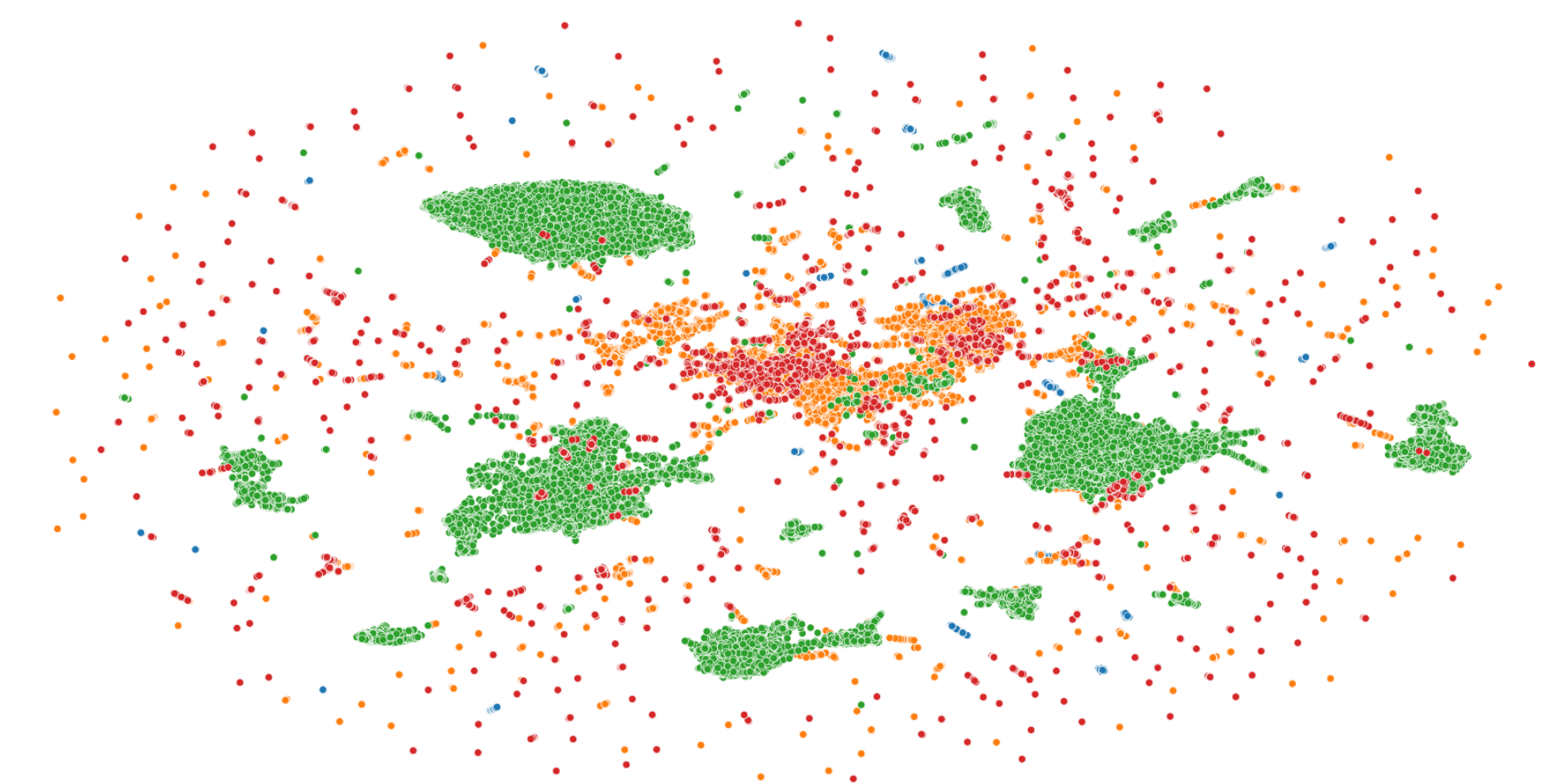
Table 1: Comparing accuracies of dual vs. multitask learning using the ResNet.

model	location	date	font	script
date and location	0.67	0.25	-	-
script and font	-	-	0.98	0.89
script and location	0.67	-	-	0.87
script and date	-	0.32	-	0.91
font and date	-	0.24	0.97	-
font and location	0.67	-	0.97	-
all tasks	0.66	0.32	0.97	0.88

Table 2: Comparing accuracies of single vs. multitask learning using the ResNet.

model	location	date	font	script
location only	0.77	-	-	-
date only	-	0.49	-	-
font only	-	-	0.98	-
script only	-	-	-	0.90
all tasks	0.66	0.32	0.97	0.88

UMAP of class distributions



Distinct clusterings

- Font (green)
- Date (orange)
- Script (red)
- Location (blue)

Perceiver Results

Table 3: Comparing accuracies of dual vs. multitask learning using the Perceiver.

model	location	date	font	script
date and location	0.60	0.19	-	-
script and font	-	-	0.96	0.80
script and location	0.65	-	-	0.78
script and date	-	0.24	-	0.80
font and date	-	0.17	0.97	-
font and location	0.58	-	0.95	-
all tasks	0.64	0.23	0.95	0.78

Table 4: Comparing accuracies of single vs. multitask learning using the Perceiver.

model	location	date	font	script
location only	0.67	-	-	-
date only	-	0.46	-	-
font only	-	-	0.96	-
script only	-	-	-	0.82
all tasks	0.64	0.23	0.95	0.78

Conclusions

- Strong competition between tasks
- Results are architecture independent
- Small semantic shifts can lead to strong incompatibility
- Multitask learning can have adversarial results when trained on shifted data-manifolds
- This may be mitigated to some degree by semantically closer tasks
- One has to be very careful in the application of multitask learning

References

- [1] Mathias Seuret et al. “ICDAR 2021 Competition on Historical Document Classification”. In: *ICDAR 2021*. 2021, pp. 618–634.
- [2] Kaiming He et al. “Deep Residual Learning for Image Recognition”. In: *CVPR 2016* (2016), pp. 770–778.
- [3] Andrew Jaegle et al. “Perceiver: General Perception with Iterative Attention”. In: *ICML*. 2021.

Contact



Alexander Mattick

Pattern Recognition Lab
Friedrich-Alexander University Erlangen-Nürnberg
Erlangen, Germany

+49 9131 85 56 58 68
alex.mattick@fau.de