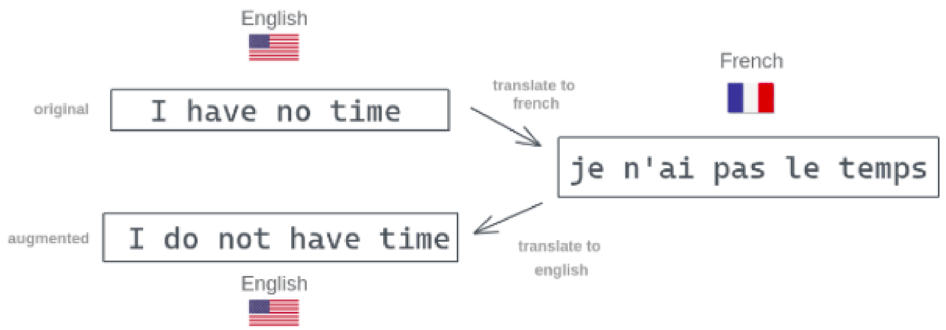
Text Is an Image: Augmentation Via Embedding Mixing

Kevin Zhao

Mentors: Vladislav Lialin and Dr. Rumshisky

Text Data Augmentation Techniques

- Data Augmentation: create new data based on existing data, without actually collecting more data
- In general, increasing the amount of data will improve performance of the model



- Synonym replacement
  - The quick brown fox jumps over the lazy brown dog.
  - The quick brown fox leaps over the lazy brown dog.
- Backtranslation

Image Data Augmentation Techniques

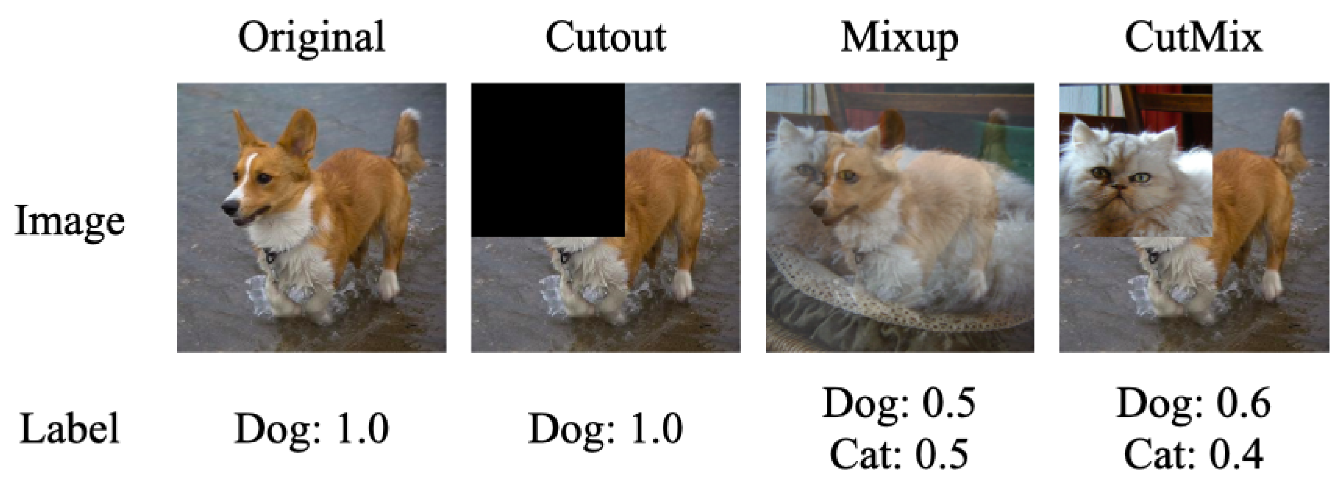
- Basic augmentation techniques
  - Crop, reflect, rotate
- Mixing based augmentation techniques
  - Cutout, Mixup, CutMix

Image Data Augmentation Techniques

$$X_{mixed} = \alpha X_1 + (1-\alpha) X_2$$
$$y_{mixed} = \alpha y_1 + (1-\alpha) y_2$$

$$X_{mixed} = M \circ X_1 + (1-M) \circ X_2$$
$$y_{mixed} = \alpha y_1 + (1-\alpha) y_2$$

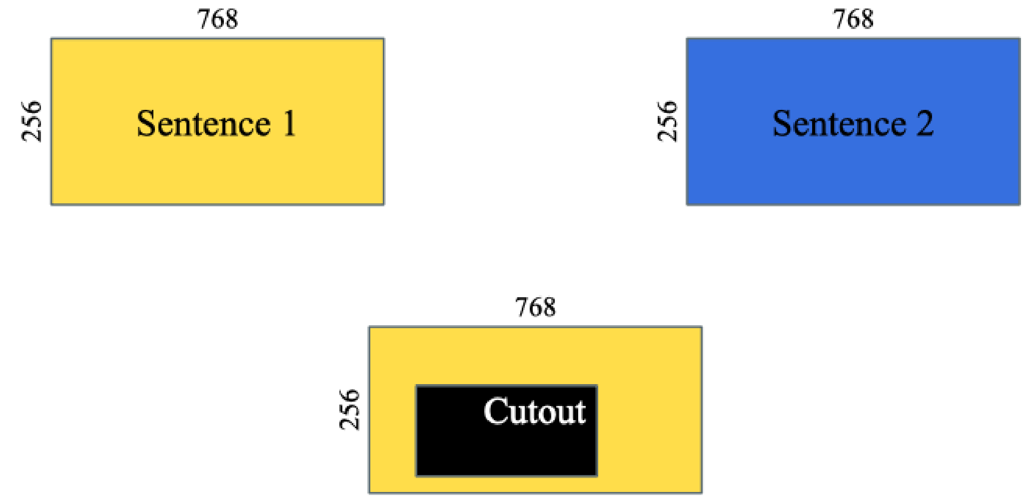
- Basic augmentation techniques
  - Crop, reflect, rotate
- Mixing based augmentation techniques
  - Cutout, Mixup, CutMix



Word Embeddings

- Pretrained
- Embedding Matrix

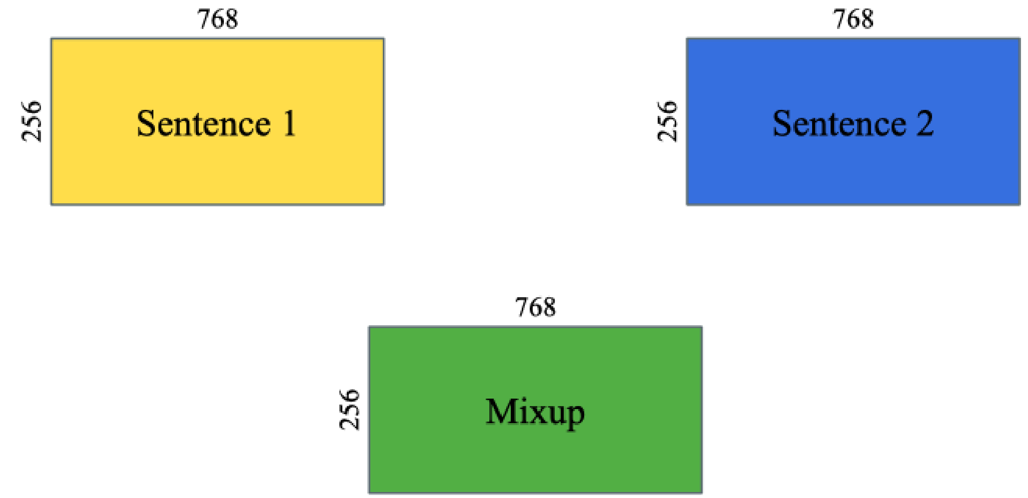
aardvark	-0.0216,	0.0426,	0.3035,	...	0.1276,	0.0043,	0.0502
abacus	0.0815,	0.1033,	-0.0632,	...	0.0426,	-0.0103,	0.0075
⋮							
hello	-0.0434,	-0.0914,	-0.1399,	...	-0.0414,	-0.1471,	0.3977
⋮							
world	0.0724,	-0.0536,	0.0984,	...	-0.0011,	0.0453,	0.0937
⋮							
zyzzyva	0.1748,	0.0247,	-0.1461,	...	-0.0119,	0.0618,	0.1744



Word Embeddings

- Pretrained
- Embedding Matrix

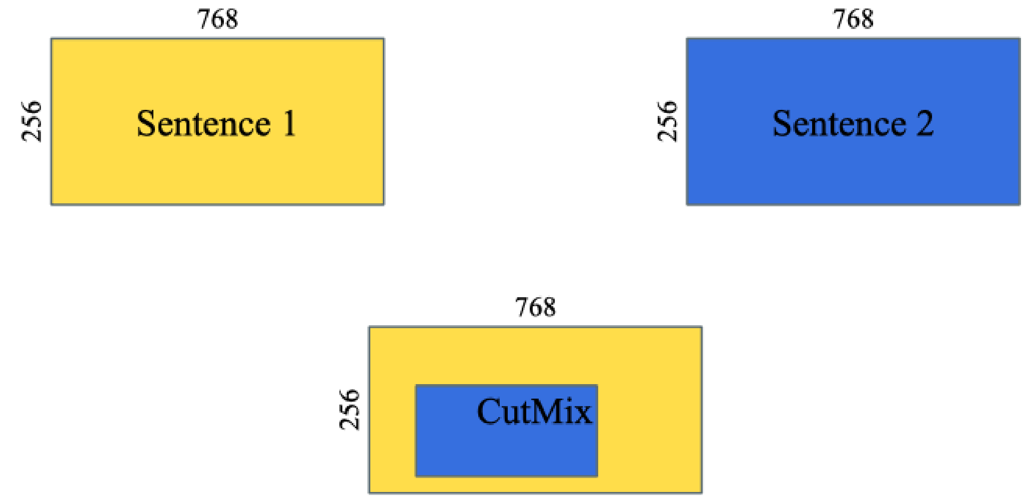
aardvark	-0.0216,	0.0426,	0.3035,	...	0.1276,	0.0043,	0.0502
abacus	0.0815,	0.1033,	-0.0632,	...	0.0426,	-0.0103,	0.0075
⋮							
hello	-0.0434,	-0.0914,	-0.1399,	...	-0.0414,	-0.1471,	0.3977
⋮							
world	0.0724,	-0.0536,	0.0984,	...	-0.0011,	0.0453,	0.0937
⋮							
zyzzyva	0.1748,	0.0247,	-0.1461,	...	-0.0119,	0.0618,	0.1744



Word Embeddings

- Pretrained
- Embedding Matrix

aardvark	-0.0216,	0.0426,	0.3035,	...	0.1276,	0.0043,	0.0502
abacus	0.0815,	0.1033,	-0.0632,	...	0.0426,	-0.0103,	0.0075
⋮							
hello	-0.0434,	-0.0914,	-0.1399,	...	-0.0414,	-0.1471,	0.3977
⋮							
world	0.0724,	-0.0536,	0.0984,	...	-0.0011,	0.0453,	0.0937
⋮							
zyzzyva	0.1748,	0.0247,	-0.1461,	...	-0.0119,	0.0618,	0.1744



Mixing Text

Original	The movie was funny and inspiring. The book had a terrible ending.
Mixup	The bookehadsafterrible endinging.
CutMix	The book was funny and inspiring.
Cutout	The movie was inspiring. had a terrible ending.
Mixup + Cutout	The terrible endinging.
Nonlinear Mixup	The bookehads funnyible endinging.

RoBERTa

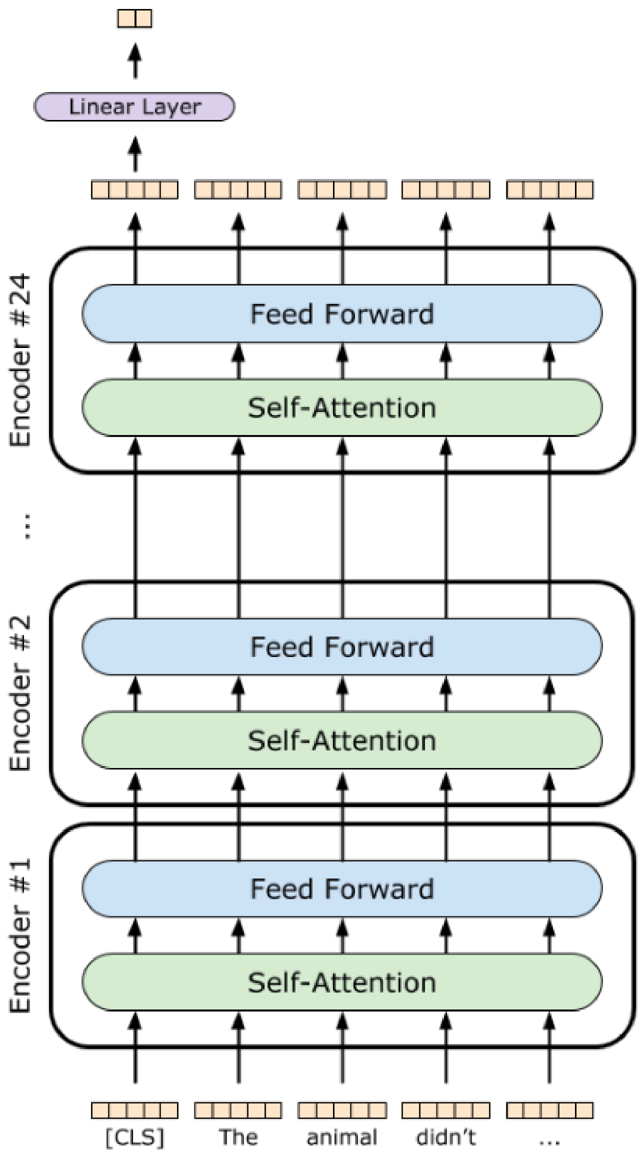
- Bidirectional
- Better long term memory
- Parallelism
- Pretrained

- Mixing on the static embeddings vs. contextualized embeddings

RoBERTa

- Bidirectional
- Better long term memory
- Parallelism
- Pretrained

- Mixing on the static embeddings vs. contextualized embeddings



GLUE

- 8 datasets

Results of augmentation techniques on GLUE tasks

RoBERTa$_{\text{large}}$	CoLA	SST-2	MRPC	STS-B	QNLI	RTE	QQP	MNLI
Reported Results	68.0	96.4	90.9/-	92.4/-	**94.7**	**86.6**	**92.2/-**	90.2
Our implementation	68.7	**96.7**	90.2/93.1	92.5/92.2	94.6	**86.6**	91.9/89.2	**90.7**
Mixup	71.2	**96.7**	90.0/93.0	92.3/92.2		79.8		
CutMix	70.7		89.7/92.7	91.8/91.5				
Cutout	67.8		89.0/92.2			81.2		
Mixup + Cutout	**72.8**	**96.7**	**92.4/94.5**	**92.6/92.5**	94.5	**86.6**	<u>92.1/89.3</u>	90.6
NonlinearMixup	65.2	96.1	90.2/93.0		94.4	**86.6**	91.2/88.3	

Conclusion

- Computer vision augmentations such as Mixup, Cutout and CutMix are applicable to natural language processing tasks when used on a word embedding level.
- Combination of Mixup and Cutout can considerably improve the results of RoBERTa on the GLUE benchmark
- A wider range of augmentations is applicable to NLP and potentially opens a way to a broader adoption of computer vision techniques in NLP

Future Work

- Nonlinear Mixup
  - AdaMixup, MetaMixup, FMix, Attentive CutMix, PuzzleMix
- Scheduler for alpha
  - No longer sample mixing weight from a fixed distribution
  - Start with weaker augmentations and progress to stronger ones

Acknowledgements

Thanks to:

- My mentors Vladislav Lialin and Dr. Rumshisky for their guidance and support
- Dr. Gerovitch and the PRIMES program for giving me this opportunity
- My parents

Questions?