4th place:
Team Kingsterdam

Team members

Phillip Lippe[1,3]

Nithin Holla[1]

Shantanu Chandra[1,4]

Santhosh Rajamanickam[1,6]

Georgios Antoniou[2,5]

Ekaterina Shutova[1]

Helen Yannakoudakis[2]

(1) University of Amsterdam

(2) King's College London

(3) QUVA Deep Vision Lab

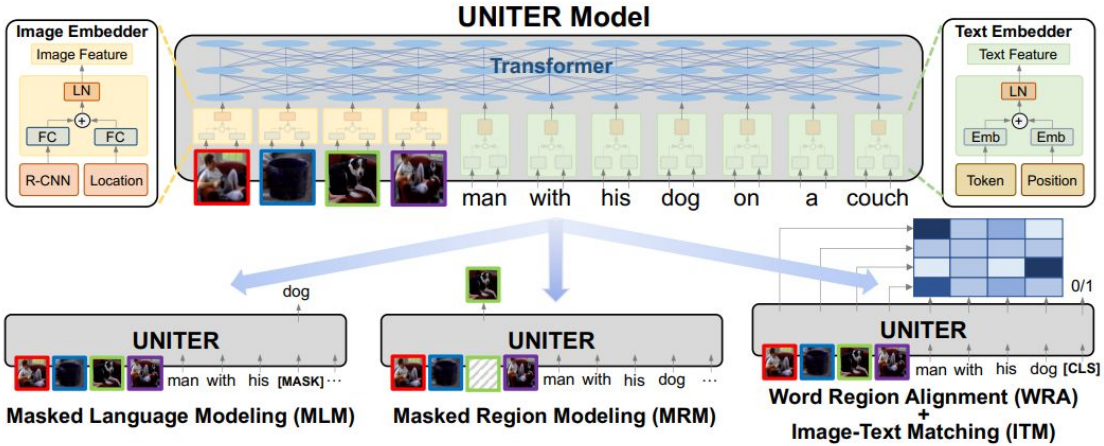
(4) ZS

(5) ClinSpec Dx™ Liquid Biopsy for Cancer

(6) slimmer AI

UNITER

- Early-fusion multimodal transformer model
- Text input: text embedding and position embedding
- Image input: bounding box features from Faster R-CNN and location features
- Pretrained on four tasks
- Other models like LXMERT and OSCAR showed considerably lower performance



Overview of UNITER[1]

[1]Chen et al. UNITER: UNiversal Image-TExt Representation learning. ECCV 2020

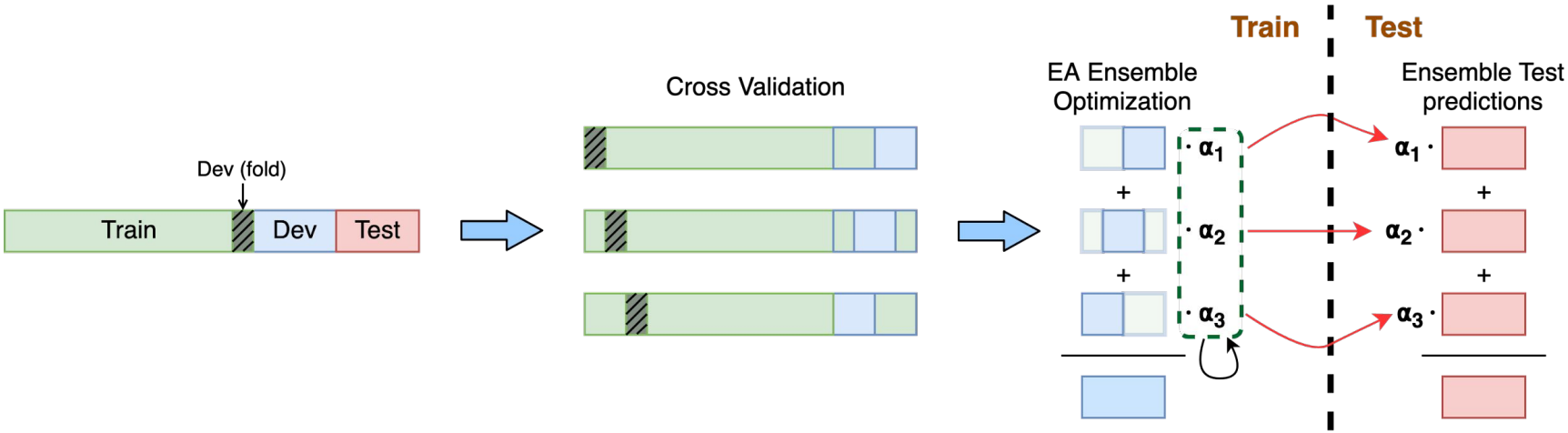
Confounders

- Benign confounders key aspect of the dataset - measures multimodality
- Simple fine-tuning performed poorly on text confounders
- Upsample text confounders during training to encourage multimodality
- Image confounders did not constitute a challenge for the pretrained model
- Class imbalance in training set ⇒ Weight loss for hateful memes higher

Example of a text confounder

Cross-validation ensemble

- Cross-validation style training on different data splits to tackle overfitting
- Development set has a larger percentage of truly multimodal examples
- Split development set for each fold into training and test
- Confounder pairs occur in the same set
- Final prediction as ensemble of 15 models, with weights optimized by an EA



Results

- Best model achieved 80.53 AUROC on Phase 2 leaderboard
- Many close runner-up models
  - CV: Simple cross-validation without including the validation in training
  - YOLO: Using fine-grained object detection of YOLO9000[1]
  - MRL: Using Margin Ranking Loss along with BCE

Model	AUROC		
	Validation	Phase 1	Phase 2
ViLBERT CC	70.07	70.03	–
Visual BERT COCO	73.97	71.41	–
$\text{UNITER}_{\text{CV}}$	80.65	79.06	–
$\text{UNITER}_{\text{MRL}}$	80.44	78.14	–
$\text{UNITER}_{\text{YOLO}}$	80.67	78.21	–
$\text{UNITER}_{\text{ENSEMBLE: CV, YOLO, MRL}}$	81.76	79.10	80.40
$\text{UNITER}_{\text{CV DEV-SET}}$	77.39	79.07	**80.53**

Table of results

[1] Redmon et al. YOLO9000: Better, Faster, Stronger. CVPR 2017

Shortcomings

- Lack of real-world knowledge
  - Fails to detect certain symbolism
  - Not aware of real-word persons

- Detecting people's characteristics
  - Not fully capable of identifying racial/religious groups
  - Does not perform well on memes suggestive of disability and abuse

Thank You!

Happy to answer questions at the Q&A session!

Team Kingsterdam: Phillip Lippe, Nithin Holla, Shantanu Chandra, Santhosh Rajamanickam, Georgios Antoniou, Ekaterina Shutova, Helen Yannakoudakis

Implementation available on GitHub: https://github.com/Nithin-Holla/meme_challenge