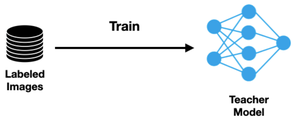
Perturb, Predict & Paraphrase: Semi-Supervised Learning using Noisy Student for Image Captioning

Arjit Jain, Pranay Reddy Samala, Preethi Jyothi, Deepak Mittal and Maneesh Singh

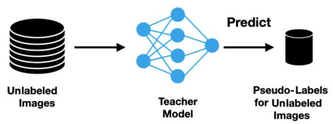
First work to use the *Noisy Student* framework for image captioning with two *key improvements*

Noisy Student

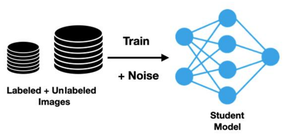
Step 1: Train teacher model on labeled data



Step 2: Use teacher model to predict pseudo-labels for unlabeled data

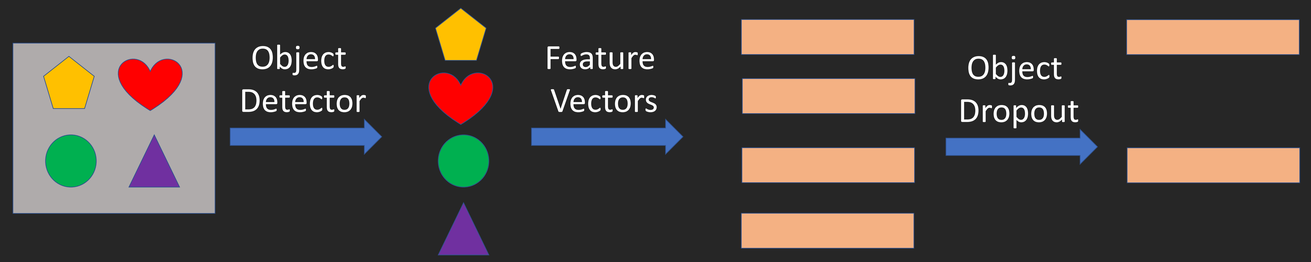


Step 3: Train a student model, with noise, using labeled and unlabeled data



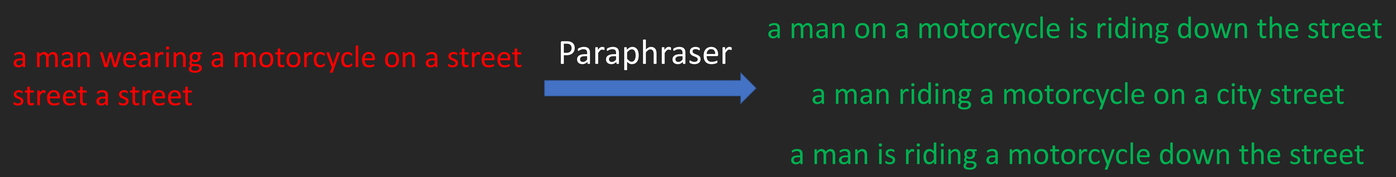
- **Training is Slow**
  Image augmentations require complete forward and backward pass through the model which is computationally expensive
- **Ignores caption-only data**
  In a low-labeled data regime, significant gains can be obtained by the clever use of caption-only data

Object Dropout



Leads to 12x faster training, with no drop in performance, by pre-computing feature vectors

Paraphraser



Provides 10% gains in performance by fixing low confidence pseudo labels instead of filtering them out

Results

Method	BLEU-4	METEOR	ROUGE-L	CIDEr	SPICE	WMD
SOTA (Unsupervised)	21.5	20.9	47.2	69.5	15.0	-
SOTA (Semi-Supervised)	25.0	21.7	49.3	73.0	14.5	16.6
Ours	**27.5**	**23.4**	**51.0**	**84.5**	**16.1**	**18.5**

Significantly outperform state-of-the-art unsupervised and semi-supervised learning methods on all metrics

Qualitative Evaluation

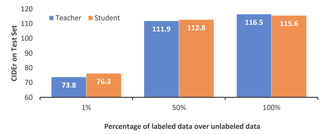
Captions generated on test images by three models:
T (teacher model), S (student model without paraphrased pseudo-labels), and SP (student trained with paraphrased pseudo labels)

T: a bear bear on a rock next to a bear bear
S: a bear bear on a rock in a rock
SP: a polar bear standing on top of a rock

T: a dog that is standing next to a stuffed bear
S: a man holding a stuffed bear on a wooden fence
SP: a group of sheep standing next to each other

Scaling Up

Keeping the unlabeled data constant, as we increase the labeled data, the gains of student over the teacher decrease



- **More unlabeled data**

Model	BLEU-4	CIDEr
Teacher	37.0	116.5
Student w 1M images	37.1	116.7

- **Weight Decay**
  Decrease the weight of pseudo-labels on the optimization objective as training progresses

Model	BLEU-4	CIDEr
Teacher	37.0	116.5
Student w weight decay	37.6	116.7

Code available at csalt-research/perturb-predict-paraphrase

IIT Bombay
Verisk Analytics