WeChat AI

Rethinking Model Pretraining for Noisy Image Classification

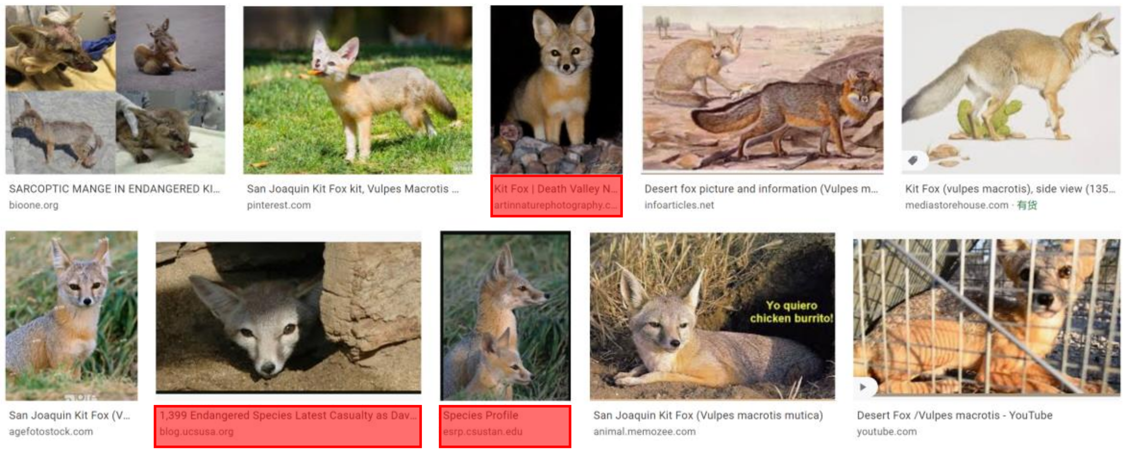
Canxiang Yan, Cheng Niu and Jie Zhou

WeChat AI

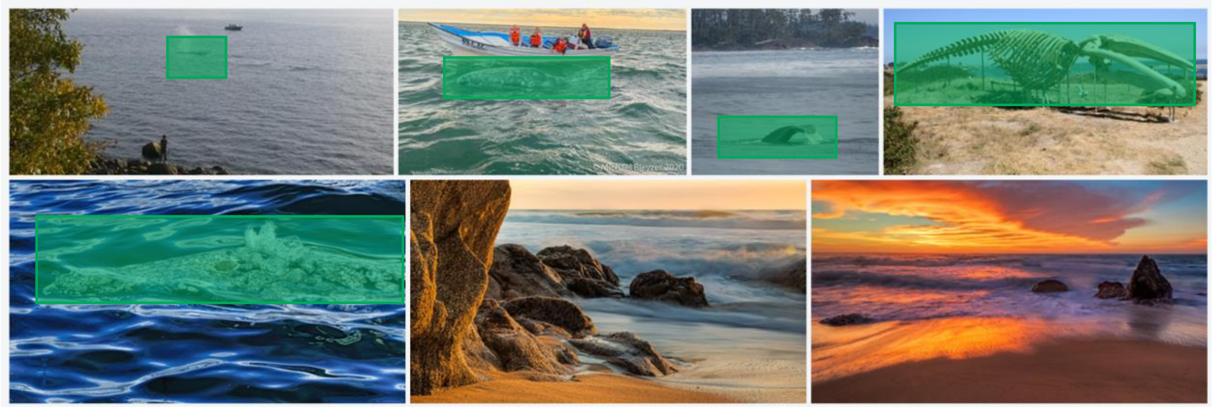
CONTENT

Noise in Webvision

- Webvision is collected from Google and Flickr
  - 5000 visual concepts and 16 million images.
  - each image may have description, title or tags.
- Noise types
  - Images with inaccurate surrounding text. → Tagging images with multiple keywords
  - Queries with unrelated reference images. → Weighting labels with semantic similarity



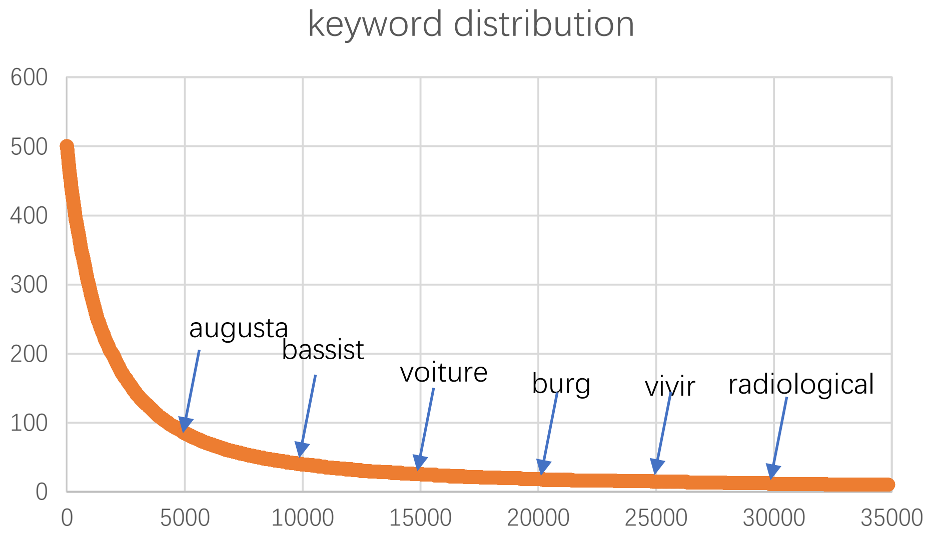
(a) Keywords missing in text. Google: Vulpes+macrotis



(b) Target missing in images. Flickr: grey+whale

Tagging images with multiple keywords

- We tag an image by extracting keywords from its context.
  - NTLK is used to recognize **nouns** and **adjectives**.
  - Most common keywords are removed, as well as least common ones.
- There are totally 35k keywords and about five for each image.



Label: n02432511 mule deer, burro deer, Odocoileus hemionus

Query: 7849 mule+deer

Description: We were hiking in the **Kaibab** National Forest south of **Williams** Arizona on the **Sycamore Rim** Trail and saw this desiccated Mountain lion scat. The mountain lion **diet** in this area consists largely of **ungulates**, more specifically **Mule deer**, **Pronghorn** and Elk. The **fur** passes through their **digestive** track and creates very **distinctive** scat. Feces of wild **carnivores** are referred to as scat. **Hunters** and **trackers** get **vital** info from scat. Because this is so desiccated, we were not in **immediate** **danger**. I've seen National Park **Rangers** diagnose the health of animals from dung and scat.

Title: Scatology 101 - Mountain lion

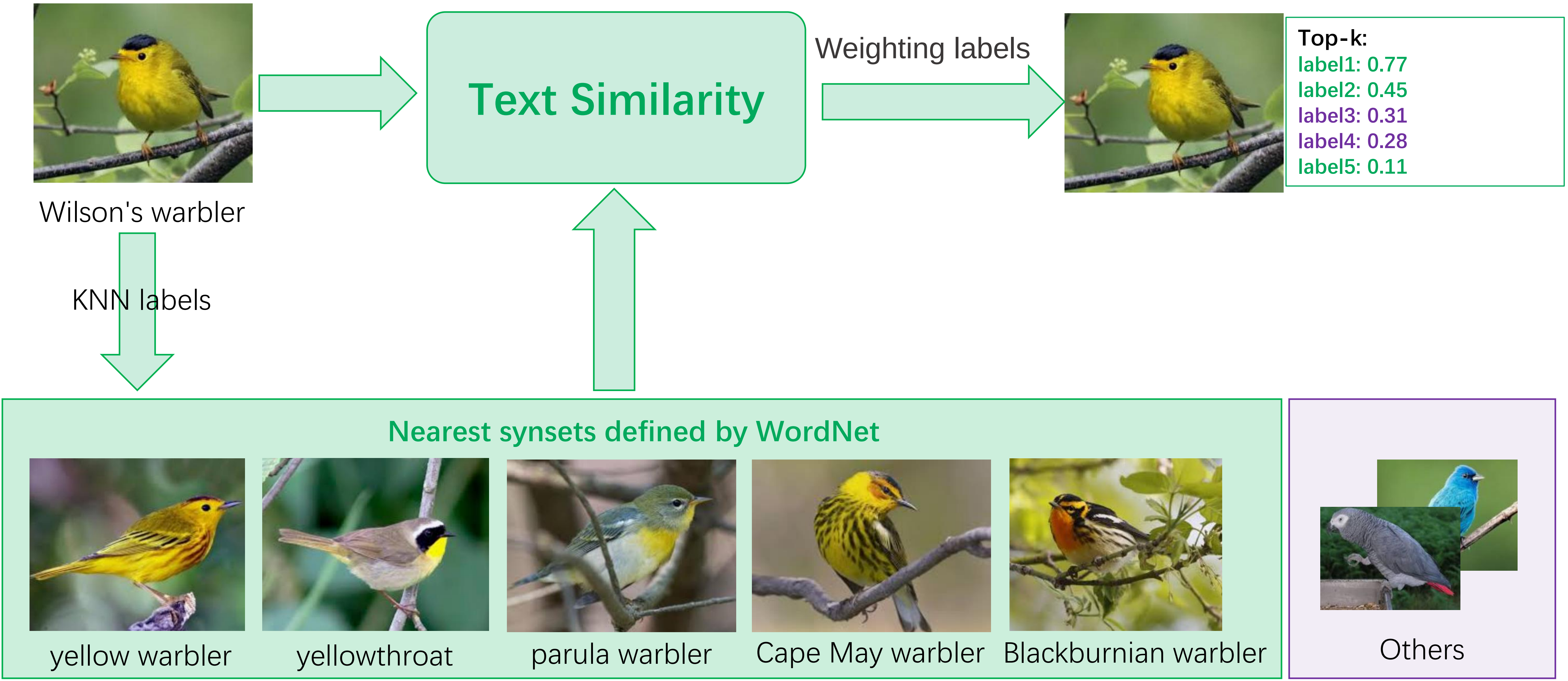
Label: n02152881 prey, quarry

Query: 9171 prey beast

Description: The cheetah examines district young **pup** **cheetah** africa savannah animal **wildcat** big cat **mammal** **mammalian** **predator** **beast** of **prey** **carnivore**

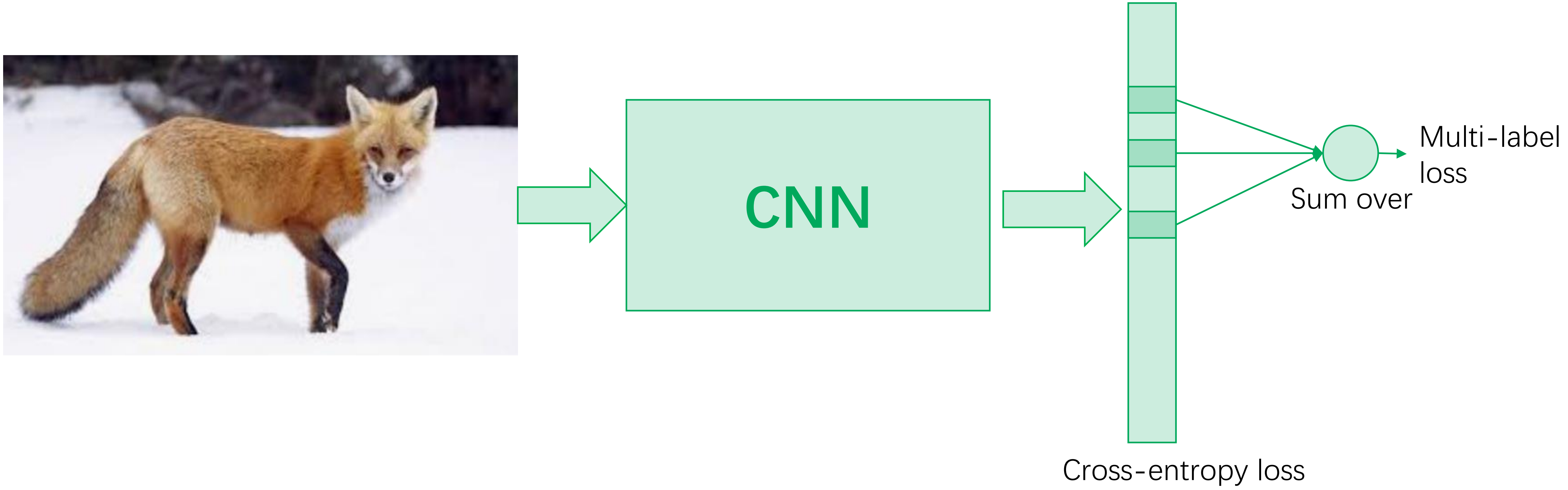
Title: **cheetah** africa savannah animal **wildcat** big cat **mammal** **mammalian**

Weighting labels with semantic similarity



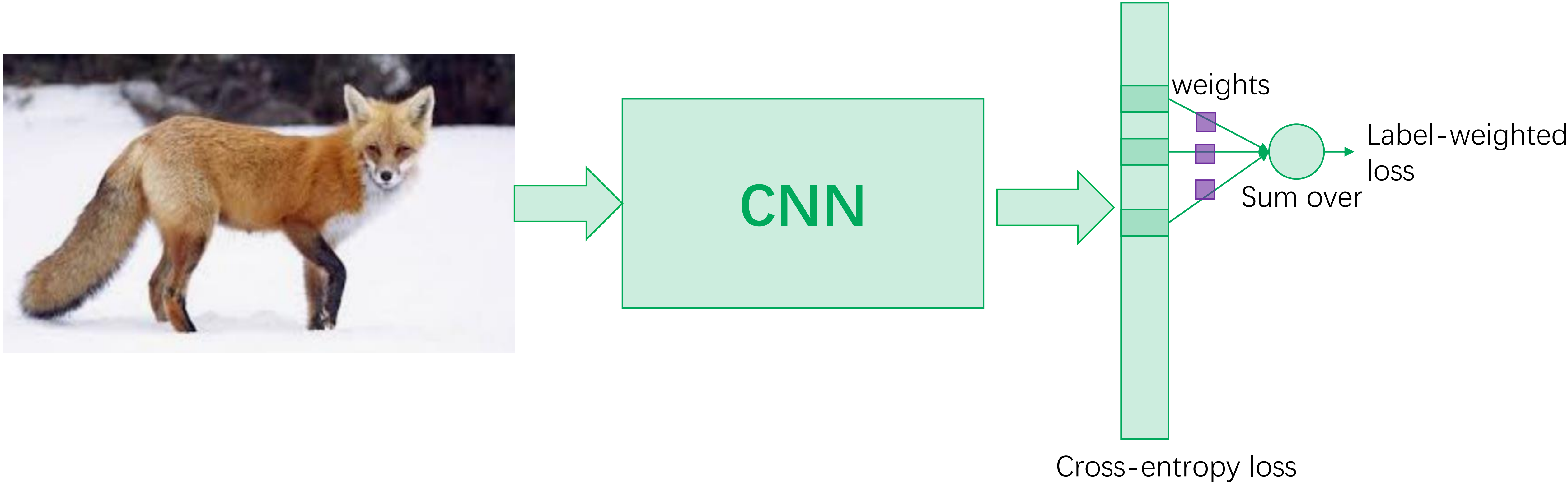
Pretraining with weakly-tagged image set (WT-Set)

- Treat it as a multi-label classification task.
- Class-balanced sampling is used for long-tail problem.
- Multi-label loss is defined to sum over cross-entropy losses on each target label.



Pretraining with label-weighted image set (LW-Set)

- Each image use weights to represent semantic correlations to the defined visual concepts.
- Based on the multi-label loss, label-weighted loss is to sum over losses with pre-defined weights on each target label.



Finetuning

- With the pretrained models on hand, we train the 5000-class model by
  - Initializing model weights except the last linear layer
  - Revising the last linear layer with 5000-dim output and random parameters.
- Dataloader:
  - Class-balanced sampling
- Optimizer:
  - SGD + Momentum
  - Learning rate: starts from 0.01, decayed by 0.1 for each 90 epochs
- Gradient Accumulation
  - Batch size: 256
  - Accumulate gradients for each 8 steps

Experiments

- Effectiveness of our pretraining

Model	Pretrain	Top1-accuaracy	Top5-accuracy
ResNeSt-101	w/o	52.0%	76.1%
ResNeSt-101	LW-Set	53.4%	76.8%
ResNeSt-101	WT-Set	55.5%	77.8%

- Different backbones

Model	Pretrain	Top1-accuaracy	Top5-accuracy
ResNeXt-101	WT-Set	55.0%	78.1%
EfficientNet-B4	WT-Set	54.4%	77.0%
ResNeSt-200	WT-Set	56.1%	78.7%

Tricks to boost performance

- Large-resolution finetuning
  - Finetune converged model with larger input size and continuous learning rate.
- Class-balanced sampling
  - It’s importance for long-tail classification
- Pseudo labeling
  - Use best models to assign pseudo labels to each image and train them again.
- Multi-model ensembling
  - Different pretraining strategies and different backbones
- Final test result

User	Entries	Date of Last Entry	top-5 accuracy ▲	top-1 accuracy ▲
fISHpAM	1	06/07/20	82.01 (2)	59.76 (2)

Conclusion

- We propose model pretraining strategies on noise images by
  - Tagging images with multiple keywords
  - Weighting labels with semantic similarity

- Experimental results prove the effectiveness of pretraining
  - Better performance
  - Faster convergence

- Future works
  - Ablation study on different keyword sets.
  - Multi-task multi-label pretraining

WeChat AI

Thanks

WeChat AI