

The Missing Link: Finding label relations across datasets Supplementary material

Jasper Uijlings, Thomas Mensink, and Vittorio Ferrari

Google Research
{jrru, mensink, vittoferrari}@google.com

A Relating ILSVRC classification and COCO segmentation

We now apply our visual ‘Pixel Probabilities’ method (Sec. 3.1) to establish relations between labels in COCO segmentation and ILSVRC12 classification. We transform our COCO segmentation model into a classification model by taking the maximum prediction score per class over all pixels in an image. We then apply it to ILSVRC12 image classification. Vice-versa, we apply a ILSVRC12 classification model to the instances in COCO by cropping images to each instance bounding box. As in Sec. 3.1, we only aggregate scores over ‘easy’ instances to establish relations.

Since for this experiment we do not have the ground truth relations, we manually inspect the top 100 relations predicted by our method. We found that 80% of them are correct. Not all relations can be found based on language alone. For example, we found that COCO’s **horse** is related to ILSVRC12’s **sorrel** (a type of horse, while ‘sorrel’ commonly refers to a plant, see Fig 1 *top row*). In 9% of the cases, the label names suggest they are in a part-of relation. However, inspecting the visual examples reveals that in some cases this is not true. For example, we predict **toilet** - **toilet_seat** to be in an identity relation. In fact, most **toilet_seats** are full toilets (only 5% **toilet_seat** with no toilet, and even 4% **toilet_seat** without seat). Another example is **potted plant** - **pot**. Again, most **pots** contain a plant, with 8% depicting only a plant, and 5% only a pot (Fig. 1 *bottom right*). Finally, we found **airplane** - **wing**, where the latter is indeed an airplane wing, not an animal wing. Here 25% of the **wing** images depict a full airplane (Fig. 1 *middle row*).

The remaining 11% of predicted relations are wrong, often due to contextual errors. For example, we predict **snow** - **ski** (Fig. 1 *bottom right*). Such mistakes may be avoidable by using COCO masks instead of boxes or by using language priors.

Finally, we investigated all **dog** - **ilsvrc** relations. Our method predicts 159 such relations, all as type *parent-of*. Remarkably, all 118 finegrained ILSVRC dog labels are included in the highest scored 129 relations we discovered.

We conclude that our method can work across different types of datasets.



Fig. 1: Examples of instances of classes in COCO (in red) and ILSVRC12 (in purple) for which we find a relation. See main text for details.

B Predict Transfer Learning Gains

In this section we investigate whether label relationships between two datasets are predictive of the gains of transfer learning. For this we correlate the performance of transfer learning to the strength of the link between labels.

We use our model trained on COCO as the source model and we use ADE20k as the target dataset. For the label relations we use the links as discovered by the *WordNet with Visual Embeddings* method. As label link strength s_b for an ADE20k label b from the COCO dataset, we aggregate the scores over all labels A in COCO for which we have established a relation by taking the mean:

$$s_b = \frac{1}{|A|} \sum_{a \in A} S_{a \rightarrow b} \quad (1)$$

Since transfer learning is most useful when the target training set is small, we fine-tune the COCO source model on 1000 images of the ADE20k training set, and then evaluate per-class Intersection-Over-Union (IoU) on the (full) validation set of ADE20k. Following [?], we measure the *gains* brought by transfer learning from COCO to ADE20k as the difference of the performance of two models:

$$gains = m_{ILSVRC12 \rightarrow COCO \rightarrow ADE20k} - m_{ILSVRC12 \rightarrow ADE20k} \quad (2)$$

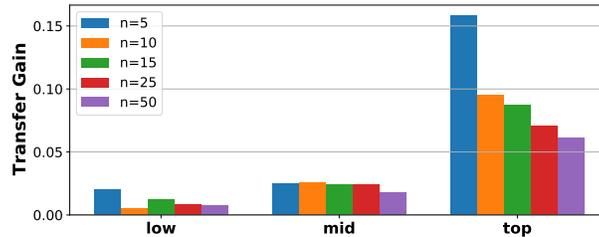


Fig. 2: Mean transfer gain over labels grouped by label strength (low, mid, high). Label strength is correlated with transfer gains.

The first model performs transfer learning from `COCO` to `ADE20k` (after initializing the `COCO` model from ILSVRC'12 as is common practice). The second model is a baseline that trains only on `ADE20k` (initialized from ILSVRC'12). This difference measures how much transferring knowledge from `COCO` helps improve performance on `ADE20k`.

In [Fig. 2](#) we show the performance gains averaged over the n labels with the weakest label link (low), the n strongest (top), and all other labels (mid). We observe that (i) the mean gain over the labels with the strongest link is higher than over the labels with the weakest link; (ii) within the top group the gain decreases as n increases, and yet it remains much higher than for middle group even for $n = 50$. This indicates that labels with a stronger label link benefit more from transfer learning than labels with a weaker relation, and that is exactly what we could have expected.

Based upon these results we conclude that transfer learning indeed does bring larger gains for target labels which have a stronger link to the source dataset.