Table 1: Evaluation of NUS-WIDE. Note that Macro/Micro P/R/F1 scores are abbreviated as O/C-P/R/F1, respectively. Ours (w/o attention) and Frequency/Rare-first (w/ atten) denote our method with the attention layer removed and using associated pre-defined label orders, respectively.

Method	C-P	C-R	C-F1	O-P	O-R	O-F1
KNN	32.6	19.3	24.3	43.9	53.4	47.6
Softmax	31.7	31.2	31.4	47.8	59.5	53.0
WARP	31.7	35.6	33.5	48.6	60.5	53.9
CNN-RNN	40.5	30.4	34.7	49.9	61.7	55.2
Resnet-baseline	46.5	47.6	47.1	61.6	68.1	64.7
Frequency-first (w/ atten)	48.9	48.7	48.8	62.1	69.4	65.5
Rare-first (w/ atten)	53.9	51.8	52.8	55.1	65.2	59.8
Ours (w/o atten)	60.8	49.5	54.5	68.3	72.4	70.2
Ours	59.4	50.7	**54.7**	69.0	71.4	**70.2**

report results on the benchmark datasets of NUS-WIDE and MS-COCO as discussed in the following subsections.

NUS-WIDE

NUS-WIDE is a web image dataset which includes 269,648 images with a total of 5,018 tags collected from Flickr. The collected images are further manually labeled into 81 concepts, including objects and scenes. We follow the setting of WARP (**?**) for experiments by removing images without any label, i.e., 150,000 images are considered for training, and the rest for testing.

We compare our result with state-of-the-art NN-based models: *WARP* (**?**) and *CNN-RNN* (**?**). We also also perform several controlled experiments: (1) removing the attention layer, and (2) fixing orders by different methods as suggested by (**?**) during training. Frequency-first indicates the labels are sorted by frequency, from high to low, and rare-first is exactly the reverse of frequency-first. The results are listed in Table 1. From this table, we see that our model performed favorably against baseline and state-of-the art multi-label classification algorithms. This demonstrates the effectiveness of our method in learning proper label ordering for sequential label prediction. Finally, our full model achieved the best performance, which further supports the exploitation of visually attended regions for improved multi-label classification.

In Fig. 3(a), we present example images with correct label prediction. We see that our model was able to predict labels depending on what it was actually attended to. For example, since ‘person’ is a frequent label in the dataset, CNN-RNN framework tended to predict it first, because their label order was defined by label occurrence frequency observed during the training stage. In contrast, our model was able to predict animal and horses first, which were actually easier to be predicted based on their visual appearance in the input image. On the other hand, examples of *incorrect* predictions are shown in Fig 3(b). It is worth pointing out that, as can be seen from these results, the prediction results were actually intuitive and reasonable, and the incorrect prediction was due to the noisy ground truth label. From the above observations, it can be successfully verified that our method is able to identify semantic ordering and visually adapt to objects with different sizes, even given noisy or incorrect label data during the training stage.

Table 2: Performance comparisons on MS-COCO. Ours (w/o attention) and Ours Frequency/Rare-first (w/ atten) denote our method with the attention layer removed and using associated pre-defined label orders, respectively.

Method	C-P	C-R	C-F1	O-P	O-R	O-F1
Softmax	59.0	57.0	58.0	60.2	62.1	61.1
WARP	59.3	52.5	55.7	59.8	61.4	60.7
CNN-RNN	66.0	55.6	60.4	69.2	66.4	**67.8**
Resnet-baseline	58.3	49.3	53.4	63.9	58.4	61.0
Frequency-first (w/ atten)	55.8	54.7	55.2	61.4	62.6	62.0
Rare-first (w/ atten)	59.5	56.5	58.0	57.3	56.7	57.0
Ours (w/o atten)	69.9	52.6	60.0	73.4	60.3	66.2
Ours	71.6	54.8	**62.1**	74.2	62.2	**67.7**

MS-COCO

MS-COCO is the dataset typically considered for image recognition, segmentation and captioning. The training set consists of 82,783 images with up to 80 annotated object labels. The test set of this experiment utilizes the validation set of MS-COCO (40,504 images), since the ground truth labels of the original test set in MS-COCO are not provided. , In the experiments, we compare our model with the *WARP* (**?**) and *CNN-RNN* (**?**) models in Table 2. It can be seen that the full version of our model achieved performance improvements over the Resnet-based baseline by 4.1% in C-F1 and by 5.6% in O-F1.

In Figures 3(c) and 3(d), we also present example images with correct and incorrect prediction. It is worth noting that, in the upper left example in Fig. 3(c), although the third attention map corresponded to the label prediction of surfboard, it did not properly focus on the object itself. Instead, it took the surrounding image regions into consideration. Combining the information provided by the hidden state, it still successfully predicted the correct label. This illustrates the ability of our model to utilize both *local* and *global* information in an image during multi-label prediction.

Conclusion

We proposed a deep learning model for multi-label classification, which consists of a visual attention model and a confidence-ranked LSTM. Unlike existing RNN-based methods requiring predetermined label orders for training, the joint learning of the above components in our proposed architecture allows us to observe proper label sequences with visually attended regions for performance guarantees. In our experiments, we provided quantitative results to support the effectiveness of our method. In addition, we also verified its robustness in label prediction, even if the training data are noisy and incorrectly annotated.

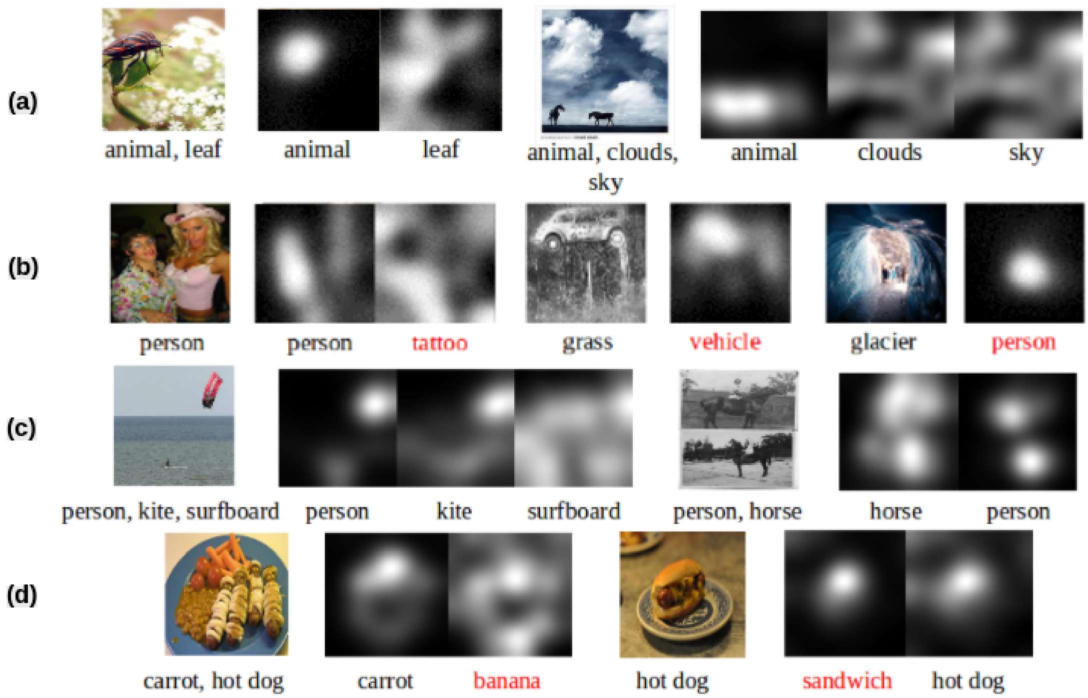

Figure 3: Examples images with correct label prediction in NUS-WISE (a) and MS-COCO (c), those with incorrect prediction are shown in (b) and (d), respectively. For each image (with ground truth labels noted below), the associated attention maps are presented at the right hand side, showing the regions of interest visually attended to. Note that some incorrect predicted labels (in red) were expected and reasonable due to noisy ground truth labels, while the resulting visual attention maps successfully highlight the attended regions.

Aspernatur quasi alias illum cum temporibus ad dolor, sint dolore ut, veniam a at vel nulla excepturi, sit cupiditate explicabo nisi veniam odio sunt delectus.Tenetur nobis ab illum esse rerum voluptates officia temporibus delectus mollitia vel, non ut ea explicabo quidem facere voluptates corrupti tenetur alias nesciunt sit.Reiciendis iure molestias itaque beatae accusamus autem ipsa similique, ut saepe officia ipsam asperiores ad laborum quam, facere quo eius tempora labore laudantium deleniti quasi voluptate a accusamus, mollitia a explicabo sequi sapiente, earum ducimus tenetur eaque cupiditate accusantium laborum?Qui repudiandae sequi nostrum, aliquam debitis distinctio sint ullam repellat accusantium facere magnam neque, tempore quibusdam similique a sed maxime distinctio non quisquam fugit eaque explicabo, nemo inventore a labore quibusdam veniam?Labore ea sint quod, eveniet dolore obcaecati quam odio esse, eaque necessitatibus numquam quas, officiis ratione minima suscipit labore nihil facere, velit similique aliquam saepe?Consequatur animi explicabo, error eos facere accusamus provident voluptate officiis quae soluta repudiandae laudantium quos?Minus architecto quam tenetur explicabo ea porro debitis, aut culpa autem asperiores enim repellat explicabo ipsam nisi tempora modi, tempore sapiente sit ullam voluptatum quas odio alias, nihil corrupti tempore sint veniam cumque natus reiciendis dolor?Officiis qui excepturi ratione rerum quod iste in exercitationem sapiente necessitatibus, cumque quibusdam iste laboriosam natus adipisci, sunt perferendis non atque quaerat quia facere, animi enim illo magnam aut modi?Recusandae assumenda perspiciatis, magni eos ducimus veritatis ab magnam id fugiat odit cumque autem, aliquid laboriosam voluptatem ipsum ad quasi magnam sint repudiandae quia, voluptates odio soluta quod libero ab esse tempora quidem deserunt, accusamus itaque id aut mollitia libero excepturi voluptatum suscipit quam magni.Delectus possimus tempora harum aperiam corporis officiis ipsam nam dolorum, debitis excepturi laboriosam totam sunt ipsum adipisci dolores eaque quo?Laborum sed et inventore, debitis voluptas nisi quaerat accusamus iusto tenetur reprehenderit nam expedita laboriosam harum, tempore aliquam sed voluptates iste quia cumque est incidunt quasi consequatur assumenda, quod nemo sequi.Quas nemo vel incidunt excepturi, autem reiciendis deserunt rerum illo vitae asperiores inventore accusantium.Iste hic saepe minima voluptate deserunt magnam, corrupti nemo atque quisquam inventore aut, consequuntur eos exercitationem corporis doloribus unde assumenda?Reiciendis at minima suscipit doloremque dignissimos, asperiores esse veritatis tempora quidem ut corporis exercitationem alias molestiae?Quam perspiciatis dolor autem placeat praesentium laboriosam laudantium id, veritatis error laudantium a, voluptatum dolorum dicta at