Towards Better Pixabay Tags

Willie Maddox

October 28, 2017

This document serves as the proposal for the final Capstone project for the Machine Learning Engineer Nanodegree offered through Udacity.

1 Domain Background

Pixabay is a website where photographers can publish and share copyright free images and videos. Since all the contents are released under the CC0 license, they are safe to use without having to ask permission or give credit to the original artist. When a user submits a new image it must first be reviewed by the Pixabay admins. They look at:

- Image Dimensions
- Focus and Blurring
- Lighting and Colors
- Copyright and Duplicates
- Image Manipulations
- Noise and JPEG Compression Artifacts
- Image Hygiene and Composition
- Tilted and Crooked Images

If the image satisfies the above categories then, most likely, it will be approved (Steinberger, 2014b). This is probably the primary reason Pixabay is so popular among photographers and artists alike; the overall image quality is professional grade.

2 Problem Statement

Along with the image upload, the user must also provide at least 3 tags describing the content of the image. The average number of tags per image is around 10. Tags make the image easily searchable by other users. Pixabay provides a tagging tutorial on their website but in general the tags are not required to meet the same level of quality standards that are placed on a newly uploaded image (Steinberger, 2012). Nor can they really be enforced. The metric for measuring the

Figure 1: *Nuts n Bolts n Boots n Pants*

quality of an image is well defined. All images must have at least 1920 along the long dimension. The image should be sharp and in focus. Avoid embedded timestamps. These are all acceptable forms of objective measurement. Tags, on the other hand, represent a persons description or interpretation of what is contained in an image and as such they are difficult to use as a source of measurement. For example, Fig.1 looks like a *nuts and bolts* to me, but to someone else it might be *hardware* or maybe even *wood*. Which tag is more *correct* is unclear. Hence, it is probably not a good idea to use tags as a measure of whether or not an image should or should not be approved.

When a user uploads an image to Pixabay, they are required to provide at least 3 labels (or tags) to describe the content of their image. Assuming that the user is the original author (or photographer) of the image, then coming up with 3 relevant tags should be trivial. However, on average, users will tend to choose around 10 tags to label their image, making it easier to find through searches.

Incorrect search tags. papillon returns lots of butterflies (no papillon tag either) Misspelled words, siberian husky not sibirian husky People end up choosing tags that are not exactly relevant. You don't have to spend a lot of time browsing pictures before you find one with a bogus label. The problem is that many of the

Can we improve the classification by adding more

types of dogs, cats, etc.

The good thing is that users who upload pictures have first hand knowledge of familiar with the content in the image and can generally be trusted to tag the image correctly. After all, an image with mislabeled tags is an image that no one will ever find. And since so much effort is required on the part of the author to get an image approved, it would seem highly unlikely that someone mislabel their own image on purpose.

When a user is first presented with the tag screen, they are asked to type in tags corresponding to the content of their image. After they type the first tag, a list of similar words (30 or so) appear for the user to select from. The list is auto-refreshed as new tags are added. This is a nice convenience that Pixabay provides, but wouldn't it be even nicer to recommend to the user in the first place a list of tags based soley on the content of the image?

In this section, clearly describe the problem that is to be solved. The problem described should be well defined and should have at least one relevant potential solution.

Additionally, describe the problem thoroughly such that it is clear that the problem is:

Quantifiable The problem can be expressed in mathematical or logical terms.

Measurable The problem can be measured by some metric and clearly observed.

Replicable The problem can be reproduced and occurs more than once. Show examples of 2-3 images that have bogus tags.

3 Datasets and Inputs

Make a table summarizing the different databases in use.

3.1 Image/Label Datasets

Imagenet images and labels COCO images and labels Pascal images and labels NUS-WIDE images and labels

3.2 Multi-Label Datasets

WordNet

- Tokenization
- Tagging - Nouns only
- Stemming
- Lemmatization
- Semantic relations - hypernym, hyponym, holonym
- Lexical relations - antonym
- Image Hygiene and Composition
- Tilted and Crooked Images

3.3 Image/Multi-label Datasets

Pixabay images and labels (i.e. tags).

Discuss the Pixabay API

The Pixabay API is well documented and it's usage is relatively straight forward (Steinberger, 2014a). At the minimum you need to pass it an API key for authentication and a query string of labels to search. For example, to retrieve web format photos about "yellow flowers", the query string q needs to be URL encoded[1]. `https://pixabay.com/api/?key=1234567-a1b2c3d4e5f6g7h8i9j0k1l2m&q=yellow+flowers&image_type=photo`. The response for this request is a json encoded data structure containing metadata for a list of images:

```
{
"total": 4692,
"totalHits": 500,
"hits": [
  {
    "id": 195893,
    "type": "photo",
    "tags": "blossom, bloom, flower",
    "webformatURL":
      "https://pixabay.com/get/..._640.jpg",
    "webformatWidth": 640,
    "webformatHeight": 360,
    "imageWidth": 4000,
    "imageHeight": 2250,
    "imageSize": 4731420,
    ...
  },
  {
    "id": 14724,
    ...
  },
  ...
]
}
```

In the response, there are three top level parameters: The `"total"` number of images in the Pixabay database with tags matching the query, the maximum `"totalHits"` that can be retrieved with the present query, and the actual `"hits"` which are a list of python dictionaries, each containing metadata about a specific image in the database. For our purposes, we only need a subset of these properties: A url to fetch the image, the set of labels that describe the image, and a mapping to link the two together. Each "hit" contains a url for the low, medium, and high resolution versions of the image. Since we will be scaling the image anyway, we can just pick the smallest one that is also larger than the input to our neural network. The medium resolution images, `"webformatURL"` don't take up a lot of space in terms of memory and download faster. The `"tags"` are the labels for the image. `"id"` to keep

---

[1] I used an invalid key here, so this url won't work. The url is maent to illustrate the basic structure of a request.

track of which image goes with which tags.

keywords described by for a list of images for use the search string to fetch All the are documented but there are restrictions that make getting the exact data you want a bit tricky. For one, when you submit a query you get back a

Search for images with Imagenet Labels

In this section, the dataset(s) and/or input(s) being considered for the project should be thoroughly described, such as:

1. how they relate to the problem
2. why they should be used
3. how the dataset or input is (was) obtained
4. the characteristics of the dataset or input

Information should be included with relevant references and citations as necessary. It should be clear how the dataset(s) or input(s) will be used in the project and whether their use is appropriate given the context of the problem (Herrera et al., 2016; Read et al., 2011; Zhang, 2006; Dave et al., 2016; Miller, 1995; Fellbaum, 1998; Zhang, Ning, and He, 2017; Loog, 2017; Wang, Yamaguchi, and Ordonez, 2017; Hu et al., 2015; Wei et al., 2014).

4   Solution Statement

Transfer learning on imagenet. Add k classes where k is the number of classes in pixabay images that are not classified by Imagenet.

For this project, I will use a pretrained model of Imagenet

The plan is to use the Imagenet model as a fixed feature extractor

1. How many images per category are there in Imagenet. (between 732 and 1300 per synset)
2. How many nouns (or physical entities) are there in WordNet.
3. How many hypernyms classes are there in Imagenet.
4. How many hyponyms per hypernym are there in Imagenet.
5. What about holonyms and meronyms. Can they be of any use with this problem?

Because the images are of such high quality on Pixabay they make great specimens for training on CNN's.

In this section, clearly describe a solution to the problem. The solution should be applicable to the project domain and appropriate for the dataset(s) or input(s) given. Additionally, describe the solution thoroughly such that it is clear that the solution is

Quantifiable The solution can be expressed in mathematical or logical terms.

Measurable The solution can be measured by some metric and clearly observed.

Replicable The solution can be reproduced and occurs more than once.

5   Benchmark Model

In this section, provide the details for a benchmark model or result that relates to the domain, problem statement, and intended solution. Ideally, the benchmark model or result contextualizes existing methods or known information in the domain and problem given, which could then be objectively compared to the solution. Describe how the benchmark model or result is measurable (can be measured by some metric and clearly observed) with thorough detail.

6   Evaluation Metrics

In this section, propose at least one evaluation metric that can be used to quantify the performance of both the benchmark model and the solution model. The evaluation metric(s) you propose should be appropriate given the context of the data, the problem statement, and the intended solution.

$$\text{IOU} = \frac{1}{N} \sum_{i=1}^{N} \frac{|y^i \wedge \hat{y}|}{|y^i \vee \hat{y}|}, \tag{1}$$

Describe how the evaluation metric(s) are derived and provide an example of their mathematical representations (if applicable). Complex evaluation metrics should be clearly defined and quantifiable (can be expressed in mathematical or logical terms).

7   Project Design

In this final section, summarize a theoretical workflow for approaching a solution given the problem. Provide thorough discussion for what strategies you may consider employing, what analysis of the data might be required before being used, or which algorithms will be considered for your implementation. The workflow and discussion that you provide should align with the qualities of the previous sections. Additionally, you are encouraged to include small visualizations, pseudocode, or diagrams to aid in describing the project design, but it is not required. The discussion should clearly outline your intended workflow of the capstone project.

Bibliography

Dave, Mihika et al. (2016). "A Novel Progressive Multilabel Classifier for Classincremental Data". In: *CoRR* abs/1609.07215. arXiv: 1609.07215. URL: http://arxiv.org/abs/1609.07215.

Fellbaum, Christiane (1998). *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press, pp. 17–31.

Herrera, Francisco et al. (2016). *Multilabel Classification*. Springer International Publishing, pp. 17–31.

Hu, H. et al. (2015). "Learning Structured Inference Neural Networks with Label Relations". In: *ArXiv e-prints*. arXiv: `1511.05616 [cs.CV]`.

Loog, M. (2017). "Supervised Classification: Quite a Brief Overview". In: *ArXiv e-prints*. arXiv: `1710.09230 [cs.LG]`.

Miller, George A. (1995). "WordNet: A Lexical Database for English". In: *Communications of the ACM* 38.11, pp. 39–41.

Read, Jesse et al. (2011). "Classifier Chains for Multi-label Classification". In: *Mach. Learn.* 85.3, pp. 333–359. ISSN: 0885-6125. DOI: `10.1007/s10994-011-5256-5`. URL: `http://dx.doi.org/10.1007/s10994-011-5256-5`.

Steinberger, Simon (2012). *Photography Training and Image Quality Standards*. `https://pixabay.com/en/blog/posts/photography-training-and-image-quality-standards-22/`. (Visited on 10/24/2017).

– (2014a). *Pixabay API*. `https://pixabay.com/api/docs/`. (Visited on 10/28/2017).

– (2014b). *Tagging Tutorial for Pixabay Images*. `https://pixabay.com/en/blog/posts/tagging-tutorial-for-pixabay-images-54/`. (Visited on 10/24/2017).

Wang, T., K. Yamaguchi, and V. Ordonez (2017). "Feedback-prop: Convolutional Neural Network Inference under Partial Evidence". In: *ArXiv e-prints*. arXiv: `1710.08049 [cs.CV]`.

Wei, Y. et al. (2014). "CNN: Single-label to Multi-label". In: *ArXiv e-prints*. arXiv: `1406.5726 [cs.CV]`.

Zhang M.L.; Zhou, Z.H. (2006). "Multi-label neural networks with applications to functional genomics and text categorization". In: *IEEE Transactions on Knowledge and Data Engineering* 18, pp. 1338–1351.

Zhang, Zhi, Guanghan Ning, and Zhihai He (2017). "Knowledge Projection for Deep Neural Networks". In: *ArXiv e-prints*. URL: `https://arxiv.org/abs/1710.09505`.