IEEE WCCI 2020

IEEE World Congress on Computational Intelligence

Virtual Conference – July 19-24, 2020

Automatic Tag Recommendation for Painting Artworks Using Diachronic Descriptions

Gianluca Zuin, Adriano Veloso, João Cândido Portinari & Nivio Ziviani

Introduction

We deal with the problem of image annotation and automatic tag recommendation for painting artworks. Diachronic descriptions containing deviations on the vocabulary used to describe each painting usually occur when the work is done by many experts over time.

To validate our method we build a model based on a weakly-supervised neural network for over 5,300 paintings with hand-labeled descriptions made in partnership by experts from the *Portinari Project*, tagging the paintings of the Brazilian painter Cândido Portinari.

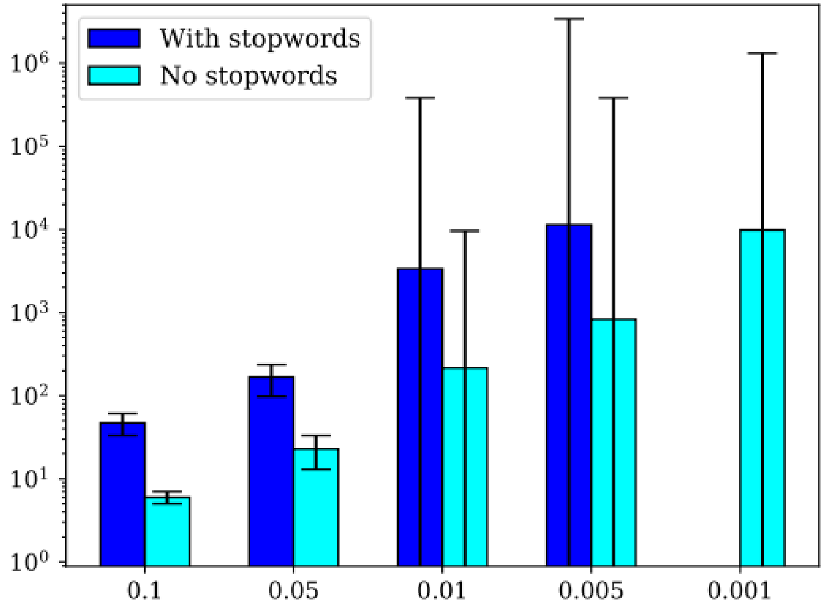
Proposed Approach

We formulate our classification model as a function $f(p,T;\theta)$ parameterized by θ that maps a painting to a set of scores associated with each candidate tag. The candidate tags (aka, classes) are extracted though itemset mining, and then paitings are classified into these classes.

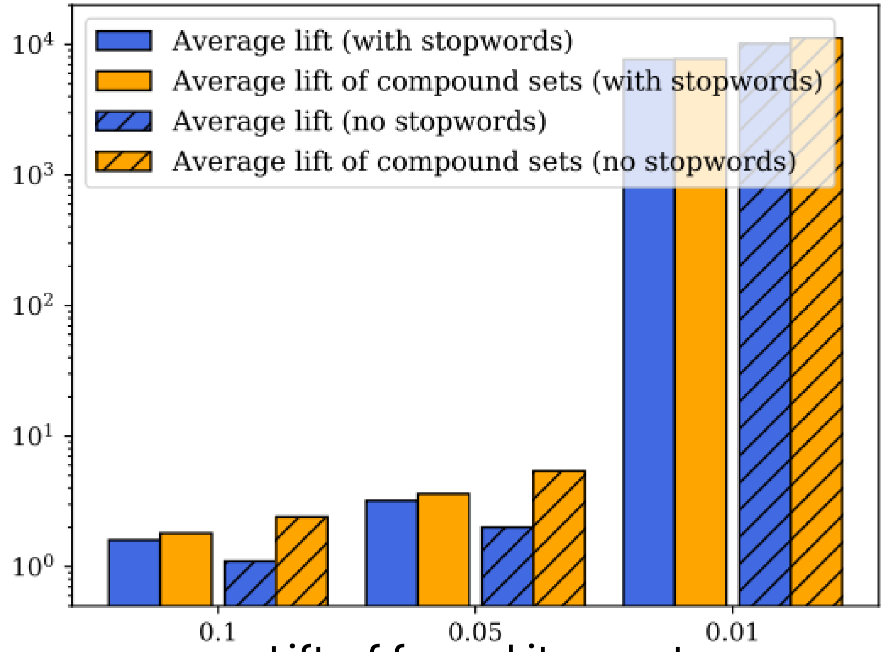
Formally, given a painting p and a list of candidate tags $T = t1,t2,...,tn$, we assume two learning scenarios: one where we learn a single $f(p,T;\theta)$ and a second in which we learn multiple $fi(p,ti;\theta)$ to calculate the relevance between p and every candidate $ti \in T$.

Itemset mining

We hypothesize that frequent itemsets provide tags with more information than simply using the most frequent words.

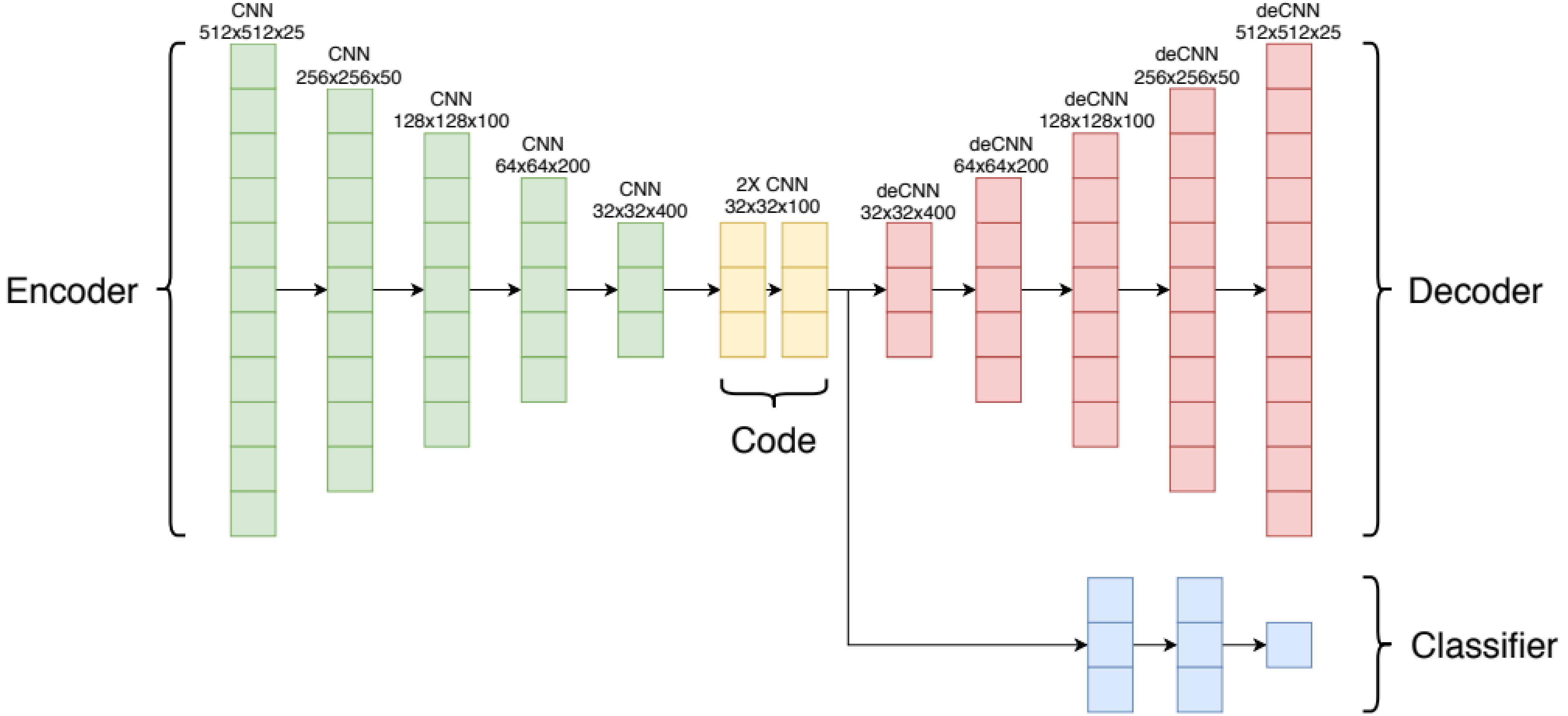


Number of itemsets found by relative support



Lift of found itemsets

Feature Learning

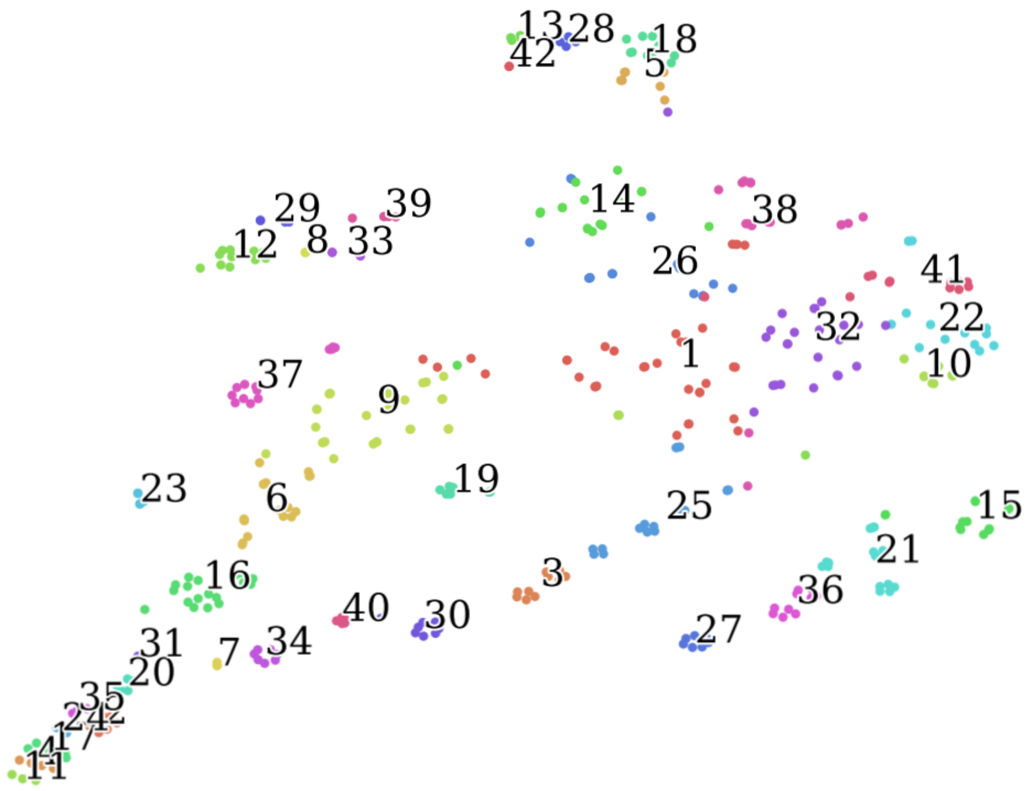


One-to-many Classifier

Models often return a superset or subset of the ground-truth labels.

They learn general painting features, which leads to tags with similar meanings being predicted in similar paintings

Solution: K-Means algorithm to cluster similar representations and retrain.



Tag-space Reduction

Tag-space	P	R	F_1	$RatioF_1$	Silhouette
100%	*.770*	*.705*	*.736*	*1.402*	-
90%	.780	.655	.712	1.389	.129
80%	.794	.631	.703	1.376	.244
70%	.798	.621	.698	1.379	.368
60%	.801	.616	.696	1.410	.495
50%	.777	.645	.705	1.487	.654
40%	.765	.691	.726	1.483	.706
30%	.797	.679	.733	1.418	.692
20%	.819	.814	.816	1.396	.601
10%	.881	.885	.883	1.379	.415

TABLE III: Autoencoder-MLP* using stopwords and mininum support set to 0.05.

Tag-space	P	R	F_1	$RatioF_1$	Silhouette
100%	*.620*	*.102*	*.175*	*1.348*	-
90%	.665	.102	.177	1.330	.084
80%	.676	.103	.179	1.334	.190
70%	.661	.114	.194	1.419	.323
60%	.673	.114	.195	1.403	.442
50%	.675	.113	.194	1.417	.493
40%	.684	.144	.238	1.460	.507
30%	.703	.184	.292	1.527	.513
20%	.724	.333	.456	1.664	.590
10%	.785	.502	.612	1.702	.604

TABLE V: Autoencoder-MLP* without stopwords and minimum support set to 0.01.

One-to-one Classifier

We training model to classify each candidate tag independently.

This allows us to perform an analysis in a case-by-case scenario, filtering undesirable models:

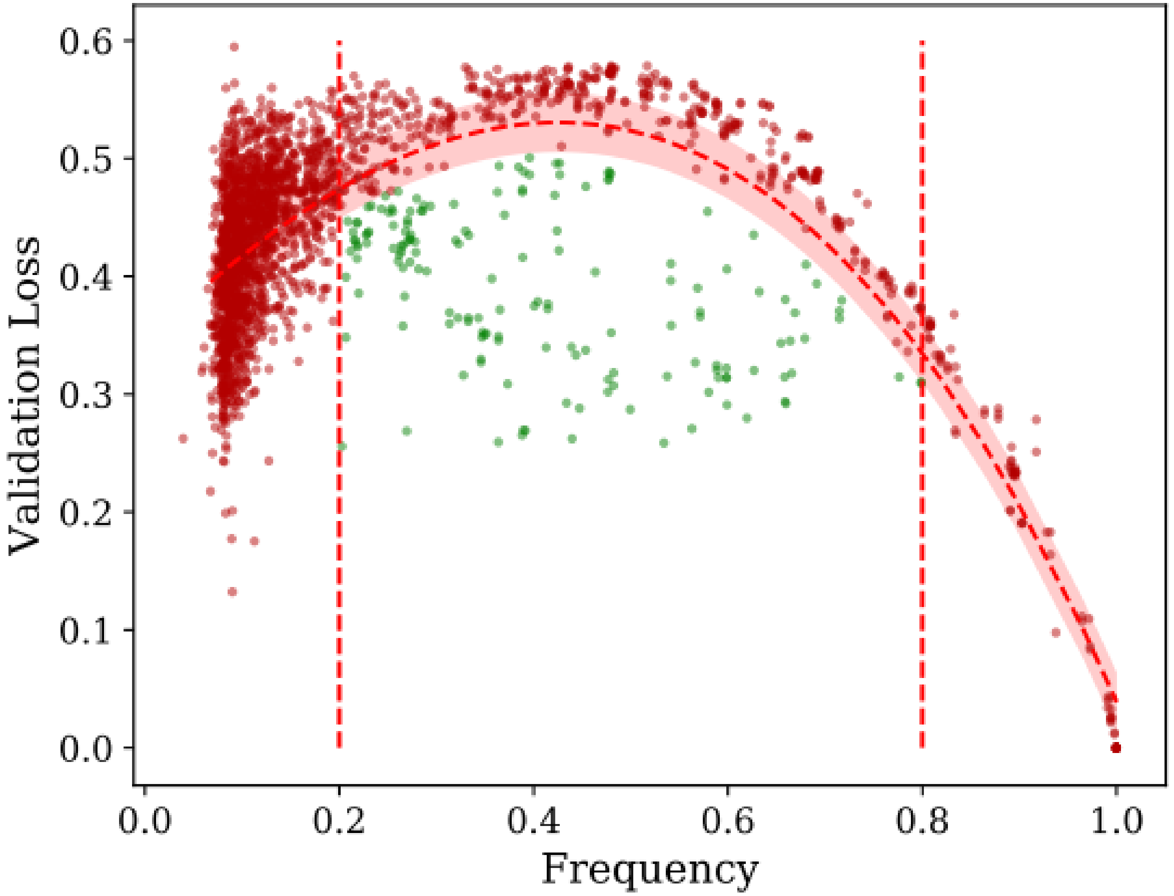
- ✗ Tags that are too *frequent* are not helpful as they do not discriminate well between paintings.
- ✗ Tags that are too *infrequent* are rare and do not provide useful information
- ✓ We wish to focus on tags that generalize well the data and provide high-performance models.

Filtering Models

The dashed lines represent the described constraints: not-frequent, not-rare and performant.

Red dots represent the models that were filtered due to being akin to random guessing given tag frequency.

Green dots represent models that learn useful features.



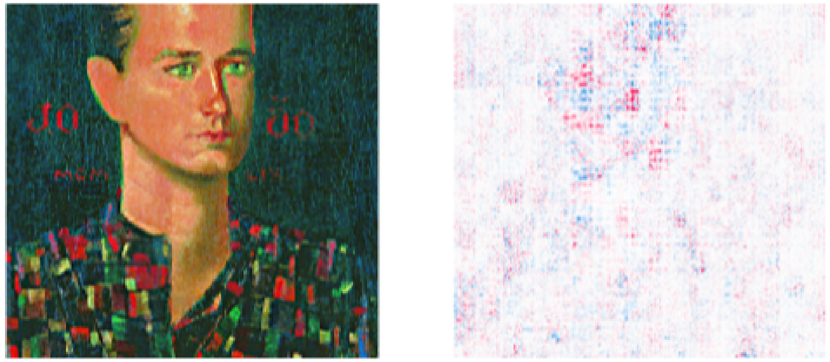
Filtering Models

Min. Support		P	R	F_1	$RatioF_1$
0.10 (with stopwords)	All tags	.805	.941	.866	1.077
	Performant	.758	.892	.818	1.119
0.05 (with stopwords)	All tags	.601	.696	.641	1.074
	Performant	.666	.826	.731	1.154
0.10 (no stopwords)	All tags	.739	.857	.789	1.075
	Performant	-	-	-	-
0.05 (no stopwords)	All tags	.350	.383	.363	1.106
	Performant	.656	.794	.715	1.409
0.01 (no stopwords)	All tags	.153	.194	.167	1.217
	Performant	.656	.794	.715	1.409

TABLE VII: Performance comparison when considering models trained on all tags or only suitable ones.

Understanding Model Output

a *"eyebrown,cheek,mouth"*



b *"cheek,mouth,nose"*

c *"cheek,mouth,nose"*

d *"area,composition"*

Conclusions

One of the main challenges of this classification task is obtaining ground-truth data for training the model. There are tags with similar semantics but that are associated with different paintings.

We proposed two modeling approaches to solve this task:

- In our first approach, we were able to achieve high performance gains while also largely reducing the tag-space size, reaching a F1 of +.85.
- On our second approach, we present a novel method for extracting meaningful tags. The tags selected are highly informative, lead to high performance and contain human-understandable explanations.

IEEE WCCI 2020

IEEE World Congress on Computational Intelligence

Virtual Conference – July 19-24, 2020

Automatic Tag Recommendation for Painting Artworks Using Diachronic Descriptions

Gianluca Zuin, Adriano Veloso, João Cândido Portinari & Nivio Ziviani