SPATIAL MORPHING KERNEL REGRESSION FOR FEATURE INTERPOLATION

Xueqing Deng, Yi Zhu and Shawn Newsam

University of California, Merced
Electrical Engineering and Computer Science
5200 N Lake Rd, Merced, CA, United States

ABSTRACT

In recent years, geotagged social media has become popular as a novel source for geographic knowledge discovery. Ground-level images and videos provide a different perspective than overhead imagery and can be applied to a range of applications such as land use mapping, activity detection, pollution mapping, etc. The sparse and uneven distribution of this data presents a problem, however, for generating dense maps. We therefore investigate the problem of spatially interpolating the high-dimensional features extracted from sparse social media to enable dense labeling using standard classification. Further, we show how prior knowledge about region boundaries can be used to improve the interpolation through spatial morphing kernel regression. We show that an interpolate-then-classify framework can produce dense map from sparse observations but that care must be taken in chosing the iterpolation method. We also show that the spatial morphing kernel improves the results.

***Index Terms*—** Feature interpolation, kernel regression, land use classification, convolutional neural network

1. INTRODUCTION

Mapping geographic phenomena on the surface of the Earth surface is an important scientific problem. Remote sensing is a traditional approach in which analysis is performed on overhead images from satellites and aircraft. This can produce dense maps but is limited by the overhead view. For example, one cannot see inside buildings.

The widespread availablility of geotagged social media has enabled novel approaches to geograhic discovery. In particular, "proximate sensing" [1] using ground-level images and videos available at sharing sites like Flickr and YouTube provides a different perspective from remote sensing, one that can see inside buildings and detect phenomena not observable from above. Proximate sensing has been applied to map land use classes [2], public sentiment [3], human activity [4], air pollution [5], and natural events [6].

A fundamental challenge in using geotagged social media to create dense maps is its sparse and uneven spatial distribution. For example, figure 1 shows the spatial distribution of Flickr images for a region of San Francisco. Even if one was able to use these images to accurately label land use, for example, the resulting map would itself be sparse and uneven.

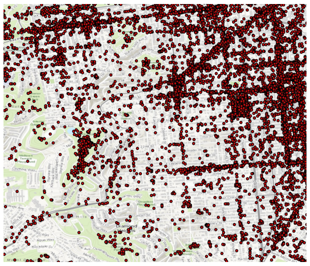

Fig. 1. Distribution of Flickr images in San Francisco. While these images can be used to map geographic phenomena such as land use, the resulting maps are sparse and uneven. We therefore investigate methods to interpolate the high-dimensional image features before performing classification.

We therefore investigate an alternate approach in which the high-dimensional features extracted from the geotagged social media are spatially interpolated before classification is performed. To our knowledge, there has been very little work on this interpolate-then-classify problem. Workman et al. in [7] spatially interpolate features extracted from street view images to match the spatial density of features extracted from overhead imagery. But, they do not investigate how best to do the interpolation. Our work in this paper performs an in-depth evaluation of the interpolate-then-classify problem using synthetic as well as real datasets.

We also investigate how to use prior knowledge about spatial heterogeneity to modulate the interpolation. We take inspiration from [8] which proposes a novel kernel that incorporates prior knowledge on spatial similarities, discontinuities, and physical and administrative boundaries to spatially interpolate a continuous variable. For example, [8] shows that knowledge of building boundaries can improve the interpolation of temperature as the indoor and outdoor temperature can be quite different. [8] only interpolates a single continuous variable and the interpolation is the final result–no classification is peformed. We instead interpolate high dimesional features extracted by convolutional neural networks (CNNs)

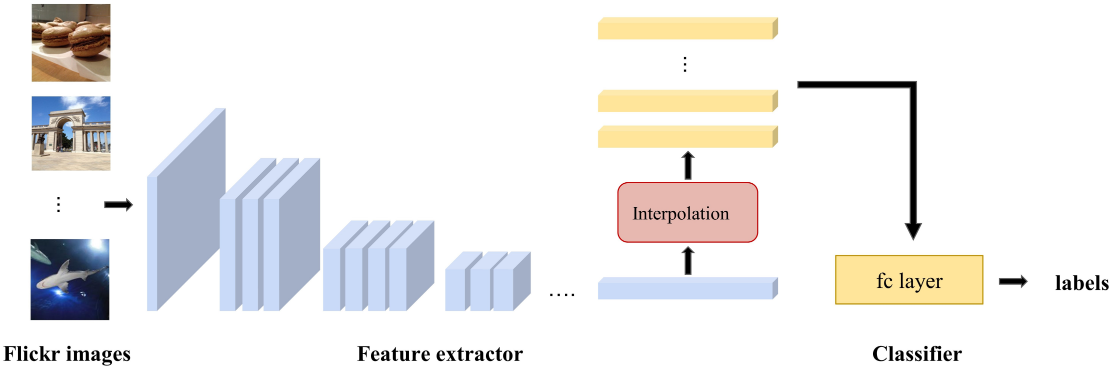

Fig. 2. Our proposed interpolate-then-classify framework. The convolutional layers of a CNN (blue) are used to perform feature extraction on sparsely located ground level images. We investigate various interpolation methods (red) including ones that incorporate prior knowledge of spatial heterogeneity. The fc layer of the CNN (yellow) is then used to perform dense classification.

with the goal of performing dense classification. The prior knowledge is incorporated through a graph Laplacian. We consider two types of graph Lapacians, one constructed using a mesh grid and another constructed using the sparse feature locations themselves.

To summarize the salient aspects of our work. We investigate the novel problem of spatially interpolating high dimensional features for dense geographic classification. We incorprate prior knowledge of spatial heterogeneity through spatial morphing kernels. And, we show results using synthetic as well as real data for mapping land use

2. METHODOLOGY

Our framework consists of three steps as shown in figure 2: feature extraction, feature interpolation, and dense classification. We use a pre-trained CNN without the final fully connected (fc) layers to perfom the feature extration. We investigate various interpolation methods including ones that incorporate prior knowledge of spatial heterogenity. Finally, the fc layers of the CNN are used to classify the densely interpolated features.

2.1. Convolutional Neural Network

Our CNN is a ResNet-101 [9] model that has been trained to label ground level images as depicting one of 45 different land use classes. (Please see [10] for more details on this model.) We separate the network into two parts: 1) a feature extractor consisting of the convolutional layers that outputs a 2,048 dimensional feature vector, and 2) a classifier consisting of the fc layer.

2.2. Interpolation

Our interpolation problem is defined as follows. Suppose we have a sparse set of n image locations $\boldsymbol{S}$=$\{s_1, s_2, ..., s_n\}$ from which we have extracted the high dimensional features $f(s_i)$. Our goal is to use these features and their locations to estimate the feature at a novel location $f(l)$. We can then create a dense feature map by densly sampling the locations l. We now describe the different interpolation methods we consider.

2.2.1. *Inverse Distant Weighting*

Inverse distance weighting (IDW) [11] is a commonly used approach to interpolate a spatially smooth surface. IDW assumes that locations that are close to one another are more alike than those that are far apart. IDW interpolation is computed as

$$f(l) = \begin{cases} \sum_{i=1}^{N} w_i(l) f(s_i), & \text{if } d(l, s_i) \neq 0 \text{ for all } i. \\ f(s_i), & \text{if } d(l, s_i) = 0 \text{ for some } i. \end{cases} \quad (1)$$

where N is the number of locations used to perform the interpolation, and $w_i(l) = 1/d(l, s_i)$ is the weight given to feature of the ith location. $d(l, s_i)$ is commonly computed as the Eulidean distance between locations l and s_i in 2D geographic space.

2.2.2. *Kernel Regression*

We also interpolate the features using Nadaraya-Waston kernel regression as is done in [7]. This interpolation is computed as

$$f(l) = \frac{\sum_{i=1}^{N} w_i(l) f(s_i)}{\sum_{j=1}^{N} w_j(l)} \quad (2)$$

Table 1. Quantitative results on toy dataset

Method	IDW	dif(%)	Gaussian	dif (%)	Spatial w/o mesh	dif(%)	Spatial w mesh	dif(%)
1	45.5	13.4	72.6	1.9	**73.2**	0.5	70.2	2.9
2	56.8	3.1	74.6	1.5	**77.2**	0.1	75.5	1.5
3	61.1	4.8	77.1	1.3	77.8	0.2	**79.8**	1.5
5	68.1	17.8	81.3	3.7	81.8	0.4	**83.9**	2.3
10	80.5	4.9	83.6	1.9	84.7	0.0	**87.4**	1.8

where $w_i(l) = k(\mathbf{x}, \mathbf{x}^{'})$ is a kernel based on the locations $\mathbf{x}$ and $\mathbf{x}^{'}$. In our case, $\mathbf{x}$

, $\mathbf{x}$ and $\mathbf{x}^{'}$ refer to the location vectors of location l and sample s_i

We use the Nadaraya-Waston regression method from [7] to interpolate features. The following formula shows that we can have different options of kernel methods. In this paper, we consider the classic Gaussian kernel function and a novel spatial morphing kernel in [8].

$$f(l) = \frac{\sum_{i=1}^{N} w_i(l) f(s_i)}{\sum_{j=1}^{N} w_j(l)} \tag{3}$$

where $w_i(l) = k(\mathbf{x}, \mathbf{x}^{'})$ is a kernel method based on the locations, $\mathbf{x}$ and $\mathbf{x}^{'}$ refer to the location vectors of location l and sample s_i

- Gaussian kernel

Let k $(\mathbf{x}, \mathbf{x}^{'})$ = $\exp(-d(l, s_i; \Sigma)^2)$, $d(l, s_i; \Sigma)$ is normalized euclidean distance, where a diagonal covariance matrix Σ controls the kernel bandwidth, and the elements are represented by a pair of trainable weights.

- Spatial morphing kernel

According to [8], having graph Laplacian to reproduce the kernel space can help incorporate spatial priori. In our paper, we have two ways to build the adjacency matrix $\mathbf{W}$. One is the same as in [8], to build adjacency matrix $\mathbf{W}$ on a dense regular mesh covering regions.$w_{ij} = 1$ if nodes i and j are connected. We refer this to spatial morphing mesh kernel(SMMK). The other is rather to use extra mesh information, we build the adjacency matrix directly of the samples based on the location boundary, which is referred to as spatial morphing sample kernel (SMSK). In this case, if two samples are in the same region then $w_{ij} = 1$. The reproducing kernel can be described as following:

$$\tilde{k}(\mathbf{x}, \mathbf{x}^{'}) = k(\mathbf{x}, \mathbf{x}^{'}) - \mathbf{k_x^T}(\mathbf{I} + \gamma \mathbf{LK})^{\mathbf{-1}} \gamma \mathbf{L} \mathbf{k_{x'}} \tag{4}$$

where $\mathbf{L}$ is the graph Laplacian, $\mathbf{L} = \mathbf{D} - \mathbf{W}$, $\mathbf{D}$ is a diagonal node degree matrix, $\mathbf{D}$ = $\sum_{i,j} w_{ij}$. Here, $\mathbf{I}$ is a unity matrix, $\mathbf{K}$=$\{k(x_i, x_j)\}_{i,j=1,...,N}$ is a kernel matrix for all data samples and $\mathbf{k_x}$ $\mathbf{k_{x'}}$ are vectors $[k(x, x_1), ..., k(x, x_N)]$ and $[k(x^{'}, x_1)$,...,$k(x^{'}, x_N)]$. Hyper parameter γ controls the spatial smoothness. A large γ means a strong enforcement on the spatial morphing.

3. EXPERIMENTS

We have two groups of experiments, first is the simulation of toy dataset, second is the real world land use classification dataset. Again, since there's no ground truth for the fine-grained land use classification which brings difficulties in evaluation. Thus we create a virtual dataset to report the quantitative and qualitative analysis of the interpolation idea. Fig 2 shows the basic pipeline of interpolating features in geographic space in our paper. To tune the kernel bandwidth, we use leave-one-out cross validation to search. And for the spatial morphing method, we set $\gamma = 100$ for all the experiments in our paper.

3.1. Simulation with toy dataset

To create a toy dataset, we first extract all the image features, and then form the class boundary and fill each pixel location with a feature vector which should be predicted with same class label within the class region. It's noted that the class boundary is not in the feature domain but in space domain. In our paper, we consider the simple three-class case. We design experiments to see how can we recover the distribution with sparse and limited examples. With the toy dataset, for each experiment we randomly select three classes out of 45 with random locations. The ground truth is 100×100 pixel, each pixel represents 1×1m space in physical world. We run the experiments for 20 times to report the average mean Intersection over Union (mIoU)[12] and the noise proportion. The noise refers to the classes which aren't included in the ground truth.

3.2. Real-world dataset: San Francisco

To apply the techniques mentioned above to address real-world problem, we extract some regions with sparse Flickr images in San Francisco to test our approach. We convert a sparse feature map to a dense feature map and then do pixel labeling.

4. RESULTS AND DISCUSSION

Fig 4 shows the visualization of some results of toy dataset. In general, we can see one class is transferring to another with

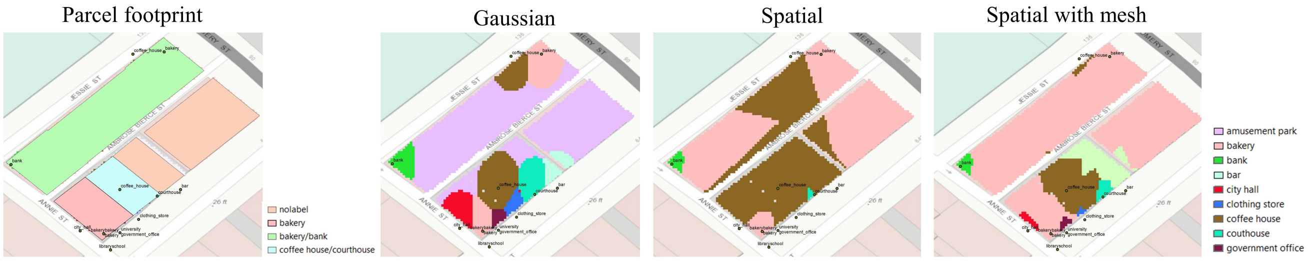

Fig. 3. Selected results on land use classification of real world dataset in San Francisco

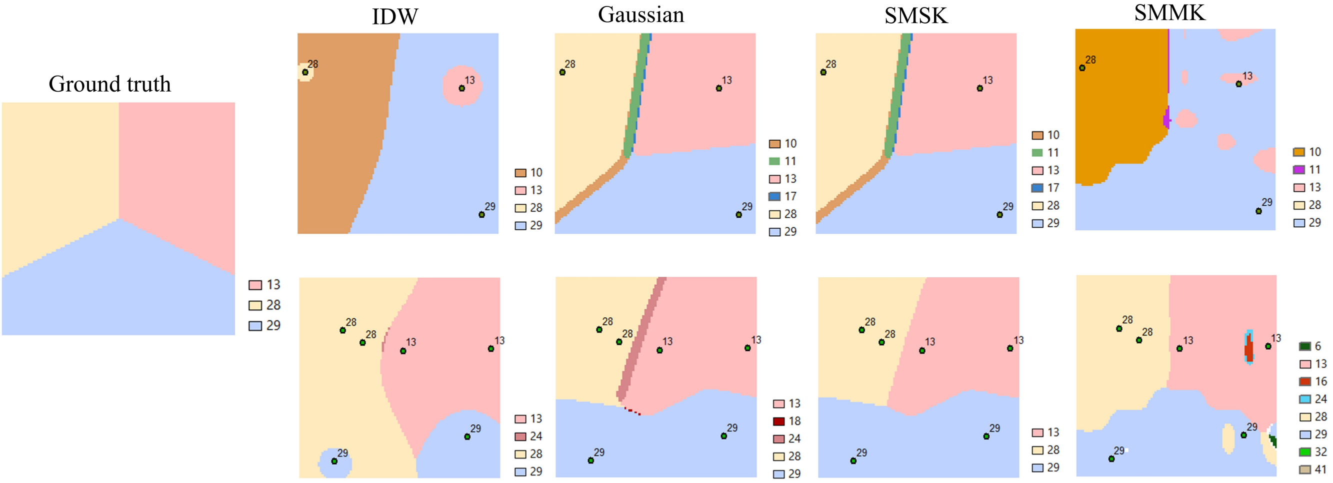

Fig. 4. Selected results on simulation of toy dataset

noisy classes on its way especially there are not enough samples to interpolate. Thus, we should be careful to interpolate features in a different domain for classification. Our approach using spatial morphing has less problem in the generating a noisy class gap interpolating features between two targets in spatial domain. Especially, for IDW with one sample to interpolate per class, the result shows a extremely wrong shape boundary. However, with kernel method, performances have great improvement. As the sample size increases, the boundary gets closer to the groundtruth. Among them, IDW has the worst result, kernel regression with Gaussian kernel and SMSK get similar boundary but there's no noisy class in SMSK. SMMK obtained the closest shape of boundary compared with ground truth.

Table 1 shows the quantitative results of the 20 simulations with different sizes of samples of toy dataset.As the mean IoU shows, IDW has trouble in interpolating features to obtain good classification accuracies. The kernel methods improve the classification accuracy dramatically. Our proposed methods work better than Gaussian kernel. When sample size increases to 10, SMMK obtained great improvement compared with the other two kernels.

Fig. 3 shows the results of the real-world dataset.In real world, one parcel may consist of multiple types of landuse, as shown in the figure, a parcel can be labeled as bakery or bank, as there are two types of landuse in one parcel. Compared with virtual dataset,the real world data has complicated boundary and difficulties in predict the class due to few samples in the region. Even though treated this land use classification problem as a pixel labeling task is challenging, SMMK result in a well-shape boundary while the others can't. And the points falling outside the shape boundary have almost no impacts on the interpolation of SMMK. However, Gaussian kernel can't hold this problem totally.

5. CONCLUSION

We have reported some observations of interpolating image features in space domain for classification problem. We found that we should be careful to interpolate high-dimensional feature vectors in two different domain. Our proposed framework can either avoid introduce noisy estimation or build a good object boundary. However, tuning the parameters of the spatial kernel with a large mesh graph Laplacian is difficult and time-consuming. In the future, we aim to improve the whole CNN architecture to convert a sparse feature map into a dense map for classification, rather than use interpolation with extracted features of CNN.

6. ACKNOWLEDGEMENT

7. REFERENCES

[1] Daniel Leung and Shawn Newsam, "Proximate sensing: Inferring what-is-where from georeferenced photo collections," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 2955–2962.

[2] Yi Zhu and Shawn Newsam, "Land use classification using convolutional neural networks applied to ground-level images," in *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 2015, p. 61.

[3] Yi Zhu and Shawn Newsam, "Spatio-temporal sentiment hotspot detection using geotagged photos," in *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 2016, p. 76.

[4] Yi Zhu, Sen Liu, and Shawn Newsam, "Large-scale mapping of human activity using geo-tagged videos," in *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, New York, NY, USA, 2017, SIGSPATIAL'17, pp. 68:1–68:4, ACM.

[5] Yuncheng Li, Jifei Huang, and Jiebo Luo, "Using user generated online photos to estimate and monitor air pollution in major cities," in *Proceedings of the 7th International Conference on Internet Multimedia Computing and Service*. ACM, 2015, p. 79.

[6] Jingya Wang, Mohammed Korayem, Saul Blanco, and David J Crandall, "Tracking natural events through social media and computer vision," in *Proceedings of the 2016 ACM on Multimedia Conference*. ACM, 2016, pp. 1097–1101.

[7] Scott Workman, Menghua Zhai, David J Crandall, and Nathan Jacobs, "A unified model for near and remote sensing," in *IEEE International Conference on Computer Vision*, 2017, vol. 7.

[8] Alexei Pozdnoukhov, "Spatial extensions to kernel methods," in *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 2010, pp. 498–501.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[10] Y. Zhu, X. Deng, and S. Newsam, "Fine-grained land use classification at the city scale using ground-level images," *ArXiv e-prints*, Feb. 2018.

[11] Donald Shepard, "A two-dimensional interpolation function for irregularly-spaced data," in *Proceedings of the 1968 23rd ACM National Conference*, New York, NY, USA, 1968, ACM '68, pp. 517–524, ACM.

[12] Jonathan Long, Evan Shelhamer, and Trevor Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.