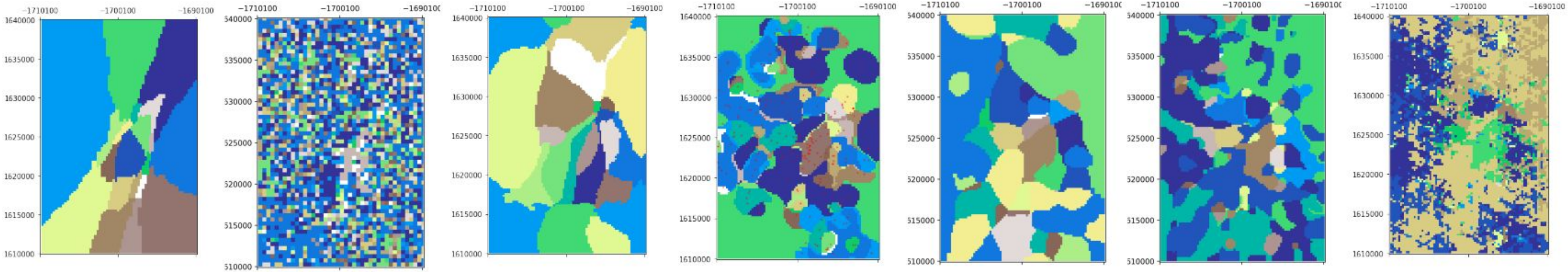
Multi-Scale Representation Learning for Spatial Feature Distributions using Grid Cells

Gengchen Mai[1], Krzysztof Janowicz[1], Bo Yan[2], Rui Zhu[1], Ling Cai[1], Ni Lao[3]
[1]STKO Lab, UC Santa Barbara; [2] LinkedIn Corporation; [3] SayMosaic Inc.



(a) *direct* (b) *tile* (c) *wrap* (d) rbf (σ=1k) (e) λ_{min}=1k (f) λ_{min}=500 (g) λ_{min}=50

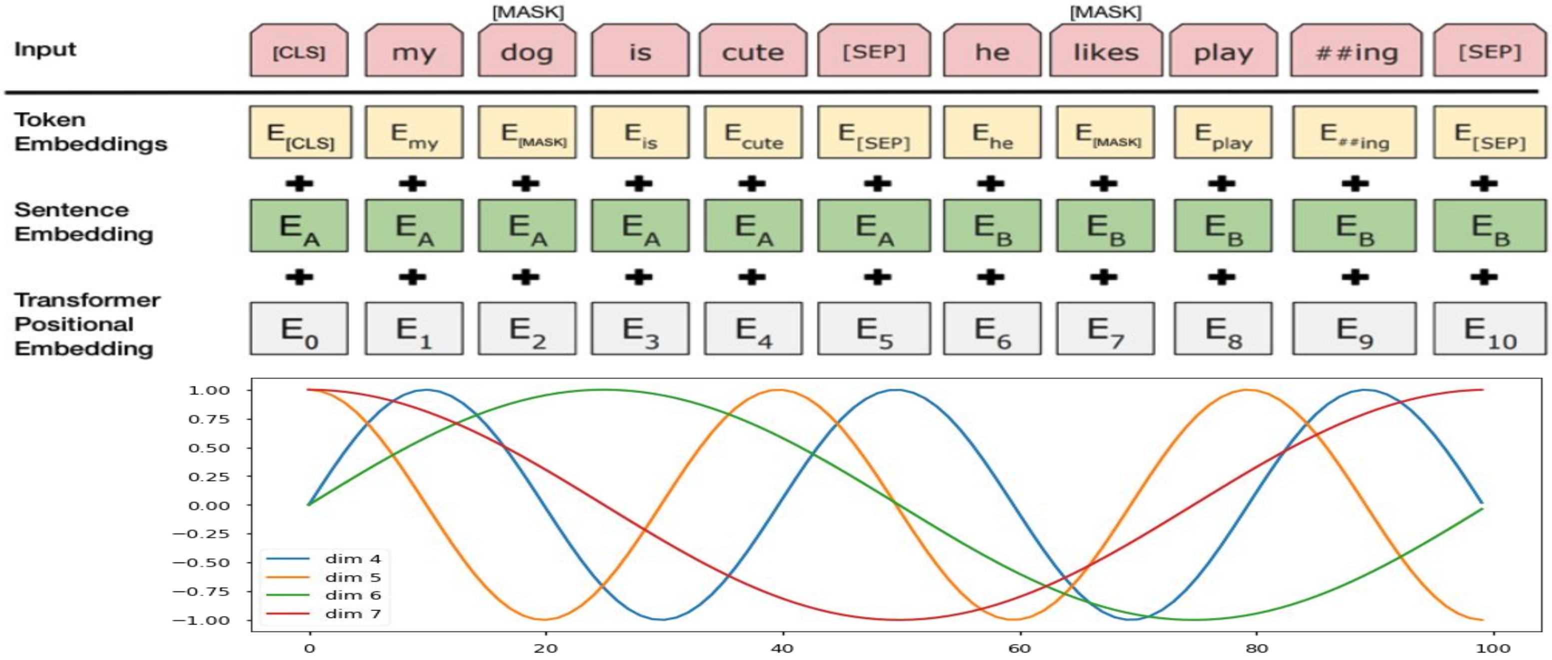
Embedding clustering of different location encoding models:
(a)-(d) baselines (e)-(f) **Space2Vec**

Arxiv paper: https://arxiv.org/abs/2003.00824

GitHub Repo: https://github.com/gengchenmai/space2vec

Unsupervised Text Encoding

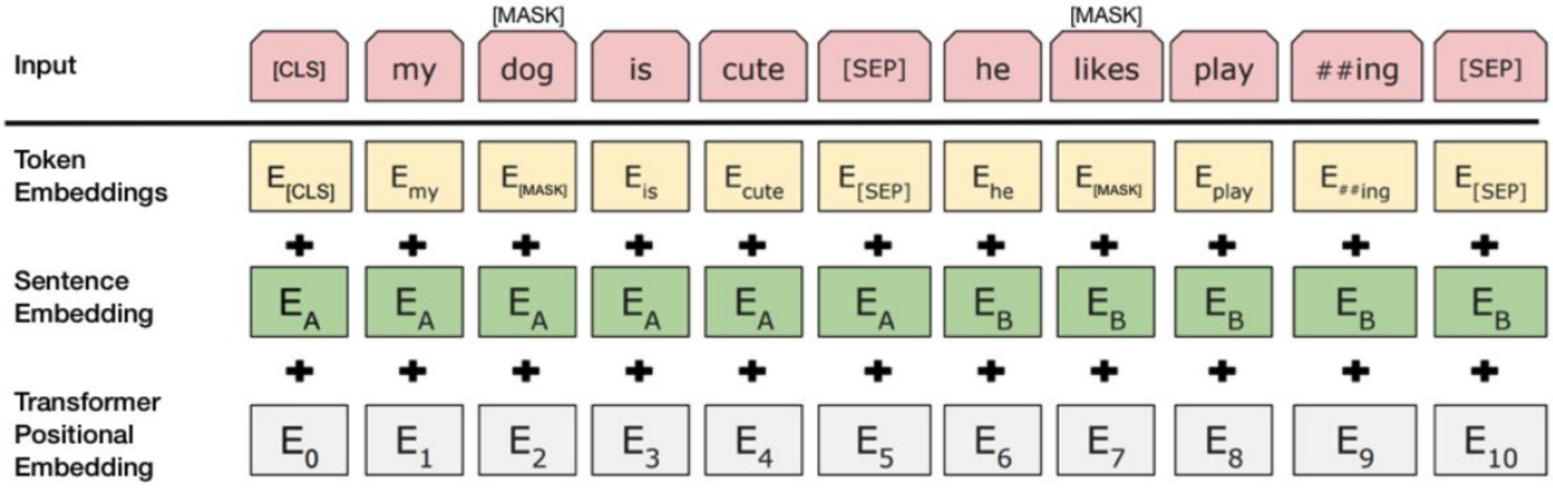
Position Encoding: encode word positions with sinusoid functions of different frequencies



Transformer (Vaswani et al., 2017) BERT (Devlin et al., 2019)

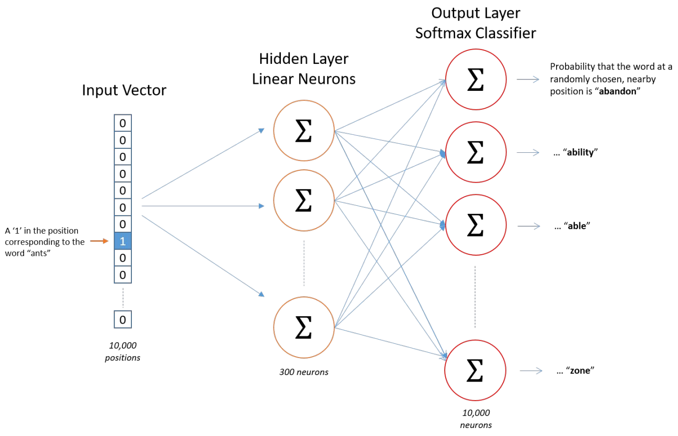
Unsupervised Text Encoding

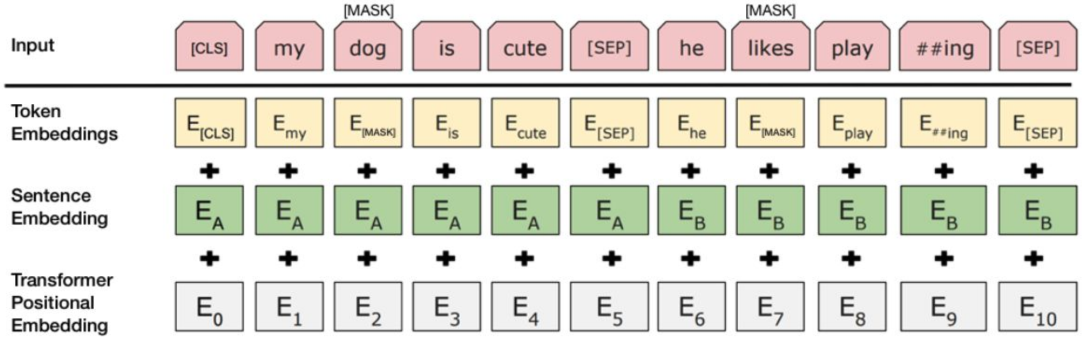
Position Encoding: encode word positions in sentences with multiple sinusoid functions with different frequencies



BERT (Devlin et al. 2019)

Unsupervised Text Encoding





Word2Vec (Mikolov et al., 2013)[1]

BERT (Devlin et al. 2019)

[1]http://mccormickml.com/2016/04/19/word2vec-tutorial-the-skip-gram-model/

Unsupervised Location Encoding

1. Radial Basis Function (RBF)

$$K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\sigma^2}\right)$$

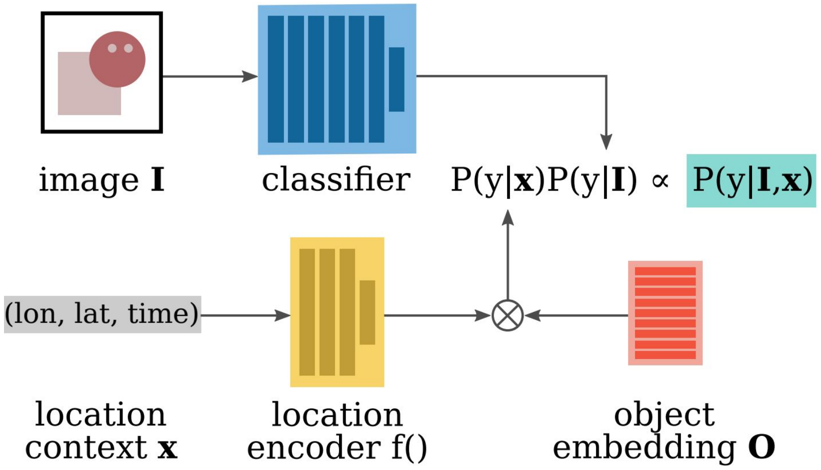
- choosing the correct scale is challenging
- Need to memorize the training samples

2. Tile-based approaches (Berg at al. 2014): discretize the study area into regular grids

- choosing the correct scale is challenging
- does not scale well in terms of memory

3. Directly feed the coordinates into a FFN (inductive single-scale location encoder)

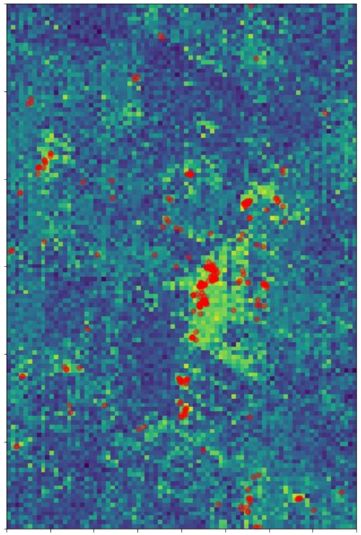
- hard to capture fine grained distributions



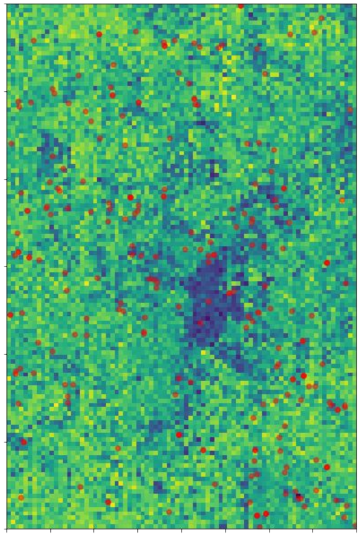
Geo-aware Image Classification (Mac Aodha et al., 2019)

Key challenge for location encoding

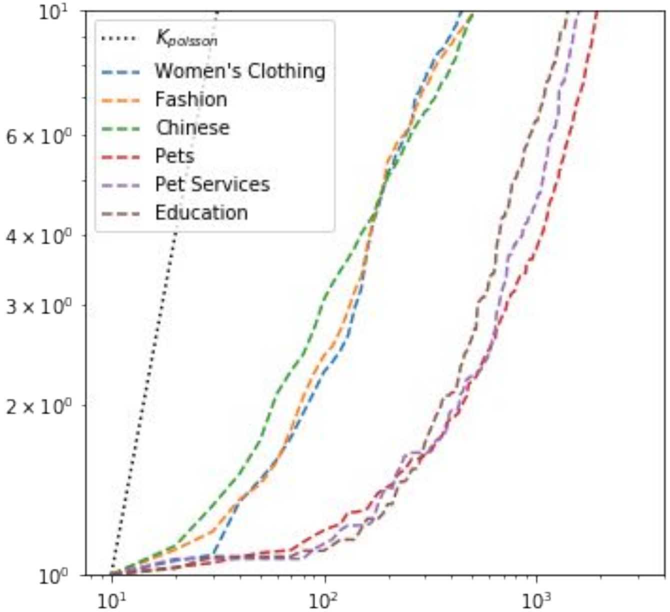
- Joint modeling distributions with very different characteristics
- => **multi-scale location representations**



Women's Clothing
(Clustered Distribution)



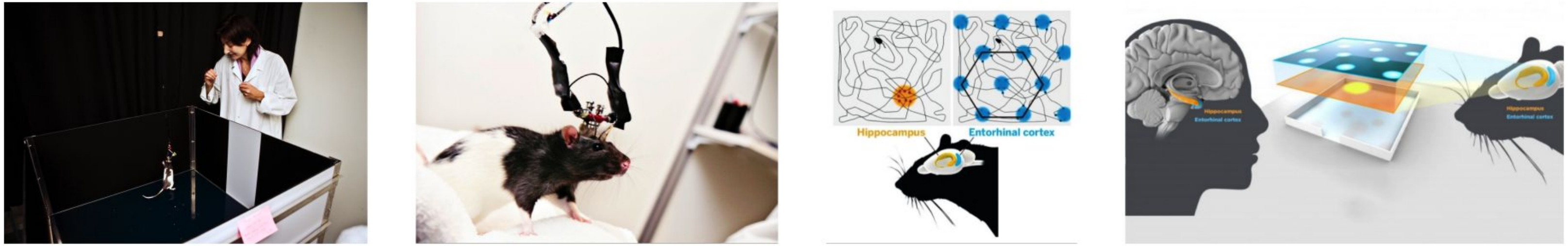
Education
(Even Distribution)



Renormalized Ripley's K
for different POI types

Stensola et al. (2012)
Gao et al. (2019)

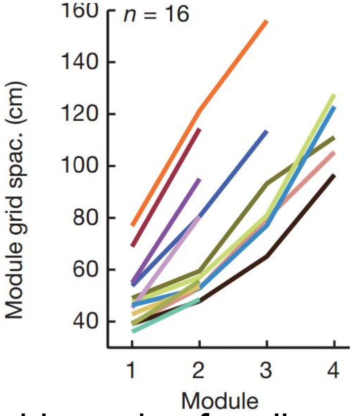
Grid Cell Based Multi-Scale Location Encoding



(a) (b) (c) (d)

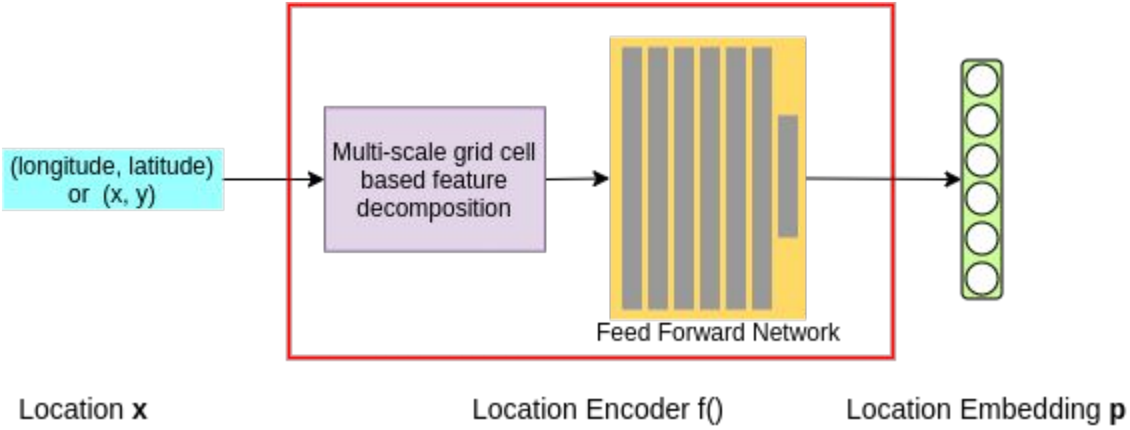
- **Grid cells** in mammals provide a **multi-scale periodic representation** that functions as a metric for location encoding.

- It can be simulated by summing **three cosine grating functions** oriented 60 degree apart (**a simple Fourier model of the hexagonal lattice**).



Mean grid spacing for all modules (M1–M4) in all animals (colour-coded)

Space2Vec



Given a location x:

$$Enc^{(x)}_{theory}(\mathbf{x}) \;=\; \mathbf{NN}(PE^{(t)}(\mathbf{x}))$$

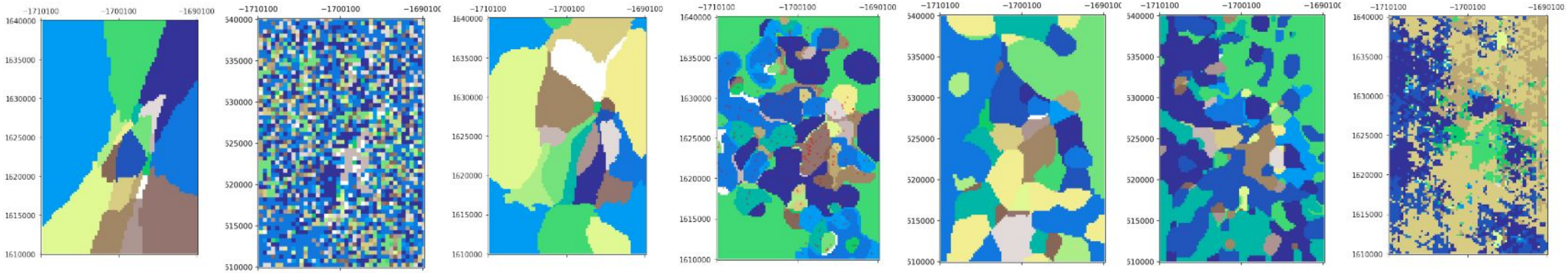
$$PE^{(t)}(\mathbf{x}) \;=\; [PE^{(t)}_0(\mathbf{x}); ...; PE^{(t)}_s(\mathbf{x}); ...; PE^{(t)}_{S-1}(\mathbf{x})]$$

$$PE^{(t)}_s(\mathbf{x}) = [PE^{(t)}_{s,1}(\mathbf{x}); PE^{(t)}_{s,2}(\mathbf{x}); PE^{(t)}_{s,3}(\mathbf{x})]$$

$$PE^{(t)}_{s,j}(\mathbf{x}) = [\cos(\frac{\langle \mathbf{x}, \mathbf{a}_j \rangle}{\lambda_{min} \cdot g^{s/(S-1)}}); \sin(\frac{\langle \mathbf{x}, \mathbf{a}_j \rangle}{\lambda_{min} \cdot g^{s/(S-1)}})] \forall j = 1, 2, 3;$$

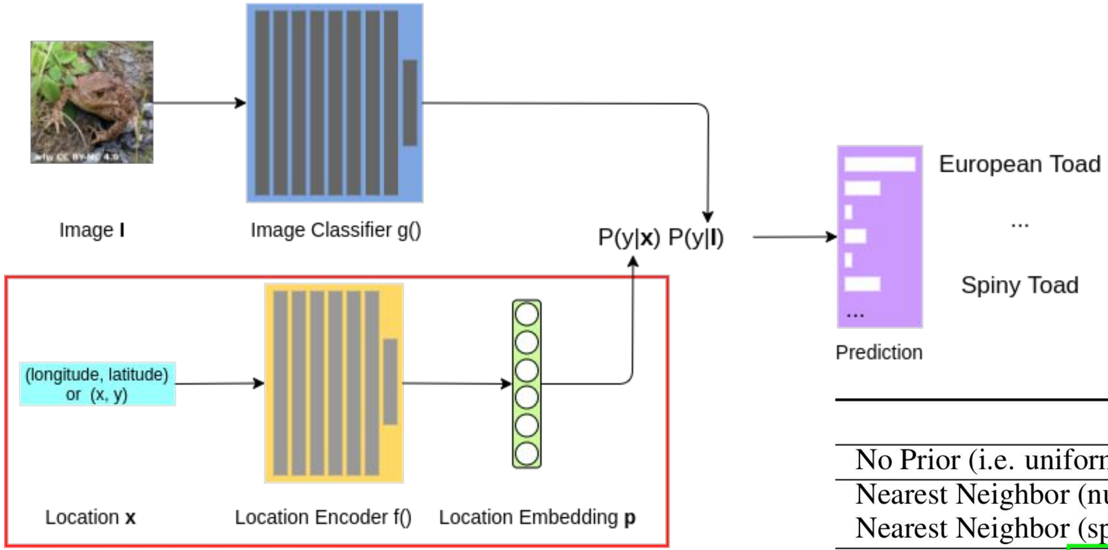
Point of Interest Type Classification

POI Groups	Clustered ($r \leqslant 100m$)	Middle ($100m < r < 200m$)	Even ($r \geqslant 200m$)
direct	0.080 (-0.047)	0.108 (-0.030)	0.084 (-0.047)
wrap	0.106 (-0.021)	0.126 (-0.012)	0.122 (-0.009)
tile	0.108 (-0.019)	0.135 (-0.003)	0.111 (-0.020)
rbf	0.112 (-0.015)	0.136 (-0.002)	0.119 (-0.012)
theory	0.127 (-)	0.138 (-)	0.131 (-)
# POI	16,016	7,443	3,915
Root Types	Restaurants; Shopping; Food; Nightlife; Automotive; Active Life; Arts & Entertainment; Financial Services	Beauty & Spas; Health & Medical; Local Services; Hotels & Travel; Professional Services; Public Services & Government	Home Services; Event Planning & Services; Pets; Education



(a) *direct* (b) *tile* (c) *wrap* (d) *rbf* (σ=1k) (e) λ_{min}=1k (f) λ_{min}=500 (g) λ_{min}=50

Geo-Aware Image Classification



(Mac Aodha et al., 2019)

	BirdSnap†	NABirds†
No Prior (i.e. uniform)	70.07	76.08
Nearest Neighbor (num)	77.76	79.99
Nearest Neighbor (spatial)	77.98	80.79
Adaptive Kernel (Berg et al., 2014)	78.65	81.11
tile (Tang et al., 2015) (location only)	77.19	79.58
wrap (Mac Aodha et al., 2019) (location only)	78.65	81.15
rbf (σ=1k)	78.56	81.13
grid (λ_{min}=0.0001, λ_{max}=360, S = 64)	**79.44**	81.28
theory (λ_{min}=0.0001, λ_{max}=360, S = 64)	79.35	**81.59**

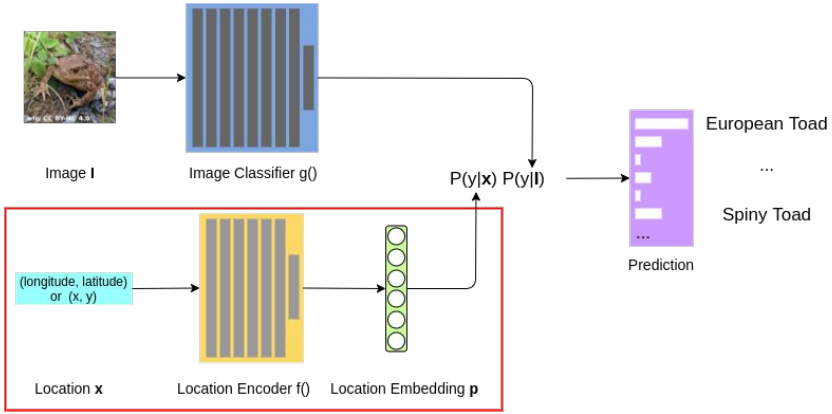
Arxiv paper: https://arxiv.org/abs/2003.00824

GitHub Repo: https://github.com/gengchenmai/space2vec

Geo-Aware Image Classification

Our **multi-scale location encoding** (*grid* and *theory*) can outperform 1) RBF (*rbf*); 2) tile-based approaches (*tile*); 3) single-scale location encoding (*wrap*).

	BirdSnap†	NABirds†
No Prior (i.e. uniform)	70.07	76.08
Nearest Neighbor (num)	77.76	79.99
Nearest Neighbor (spatial)	77.98	80.79
Adaptive Kernel (Berg et al., 2014)	78.65	81.11
tile (Tang et al., 2015) (location only)	77.19	79.58
wrap (Mac Aodha et al., 2019) (location only)	78.65	81.15
rbf (σ=1k)	78.56	81.13
grid (λ_{min}=0.0001, λ_{max}=360, S = 64)	**79.44**	81.28
theory (λ_{min}=0.0001, λ_{max}=360, S = 64)	79.35	**81.59**



Arxiv paper: https://arxiv.org/abs/2003.00824

GitHub Repo: https://github.com/gengchenmai/space2vec