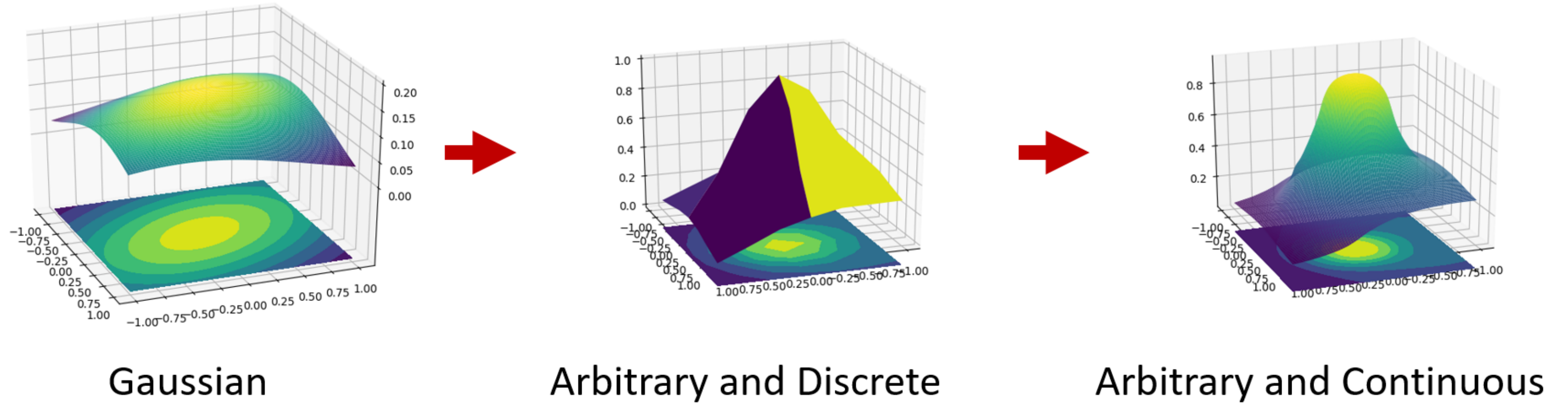
Continuous Non-Parametric Potential Estimation in Conditional Random Fields

Shi MAO, Yinheng ZHU, Lu FANG, Ercan E. KURUOGLU

Tsinghua-Berkeley Shenzhen Institute

{maos19,zhuyh19}@mails.tsinghua.edu.cn, {fanglu, kuruoglu}@sz.tsinghua.edu.cn

Introduction


Gaussian Arbitrary and Discrete Arbitrary and Continuous

- Conditional Random Fields (CRF) have been widely used for solving “neighboring dependency problem” in Computer Vision.
- The Gaussian assumption of potential function is used without validation.
- We introduce a non-parametric discrete and continuous representation of potential function which can be optimized in a data-driven manner by gradient decent.
- The learned potential shows semantic relations among categories. It can benefit domain specific image segmentation and also model diagnosis.

Discrete Method

- From Gaussian Kernel to arbitrary representation for pairwise potential [1]

$$\Psi^{p}_{f_i f_j}(x_i, x_j) = K^{b}_{x_i x_j}(f_i - f_j) + K^{s}_{x_i x_j}(p_i - p_j) \tag{1}$$

where p_i and f_i denotes the spatial (a 2D vector of location) and bilateral (a 5D vector of location and RGB values) feature of pixel i. The "neighboring dependency" is featured by these features.

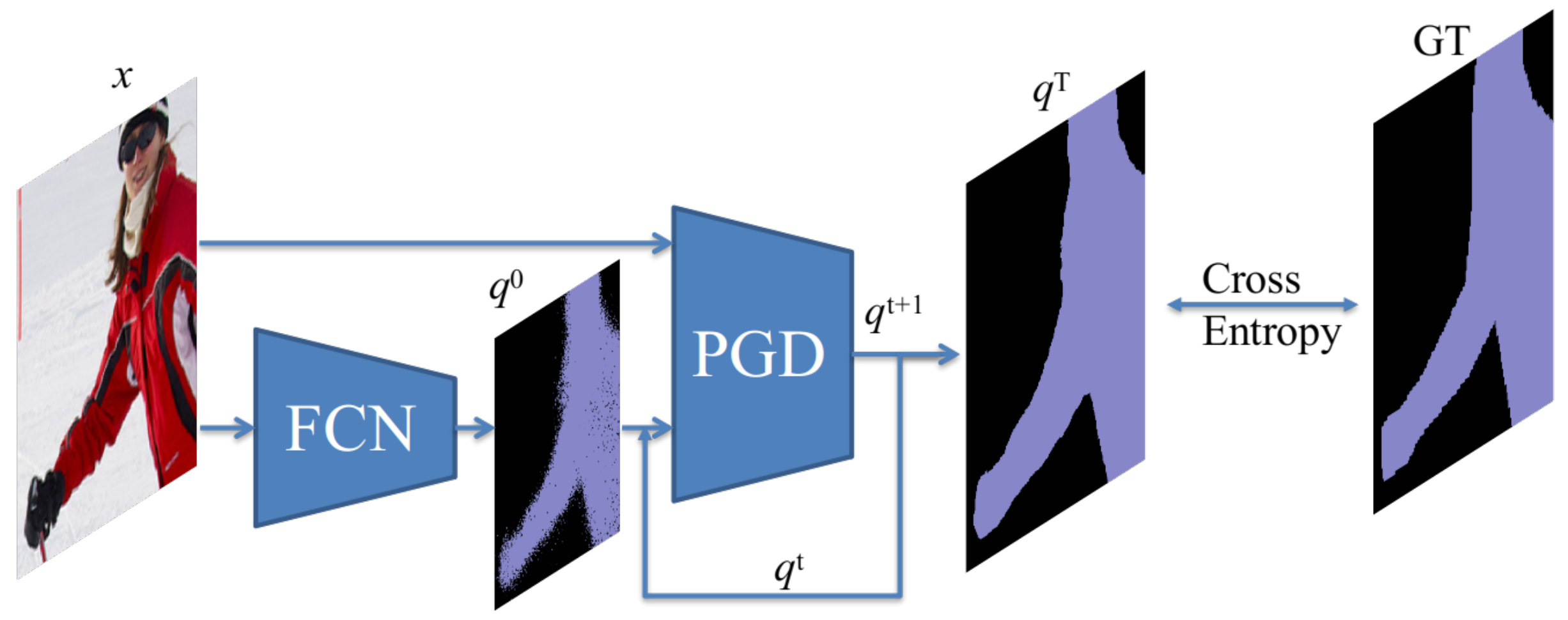
- Define differentiable energy function as a superposition of unary and pairwise potential functions of the label probabilities. Above representation for pairwise potential is used.

$$\min E(q) = \sum_{i\in\mathcal{V},\lambda\in\mathcal{L}} \Psi^{u}_{p_i}(\lambda) q_{i\lambda} + \sum_{\substack{(i,j)\in\varepsilon \\ \lambda,\mu\in\mathcal{L}}} \Psi^{p}_{f_i f_j}(x_i, x_j) q_{i\lambda} q_{j\mu} \tag{2}$$

- Minimize energy function by Projected Gradient Descent, where pairwise gradient can be calculated as convolutional operation

$$\frac{\partial E(q)}{q_{i\lambda}} = -w_u \log q^{0}_{i\lambda} + \sum_{\substack{j:(i,j)\in\varepsilon \\ \mu\in\mathcal{L}}} K^{b}_{\lambda,\mu}(f_i - f_j) q_{j\mu} + \sum_{\substack{j:(i,j)\in\varepsilon \\ \mu\in\mathcal{L}}} K^{s}_{\lambda,\mu}(p_i - p_j) q_{j\mu} \tag{3}$$

- Pipeline as a RNN to learn kernel weights



Continuous Method

As shown in Fig. 1, discrete representation is usually regarded and implemented as matrix with size of 7 × 7, 9 × 9 or 11 × 11. Without losing generality, let’s set size = 7 × 7 and the potential function $\Psi(x, y)$ is only defined on $x, y \in \{0, 1, 2, 3, 4, 5, 6\}$. It can be obtained by simply taking scalar in the corresponding location in the matrix. However the discretization of $\Psi(\cdot)$ can be a oversimplified assumption, which limits the spatial resolution and loses the local detail of potential function.

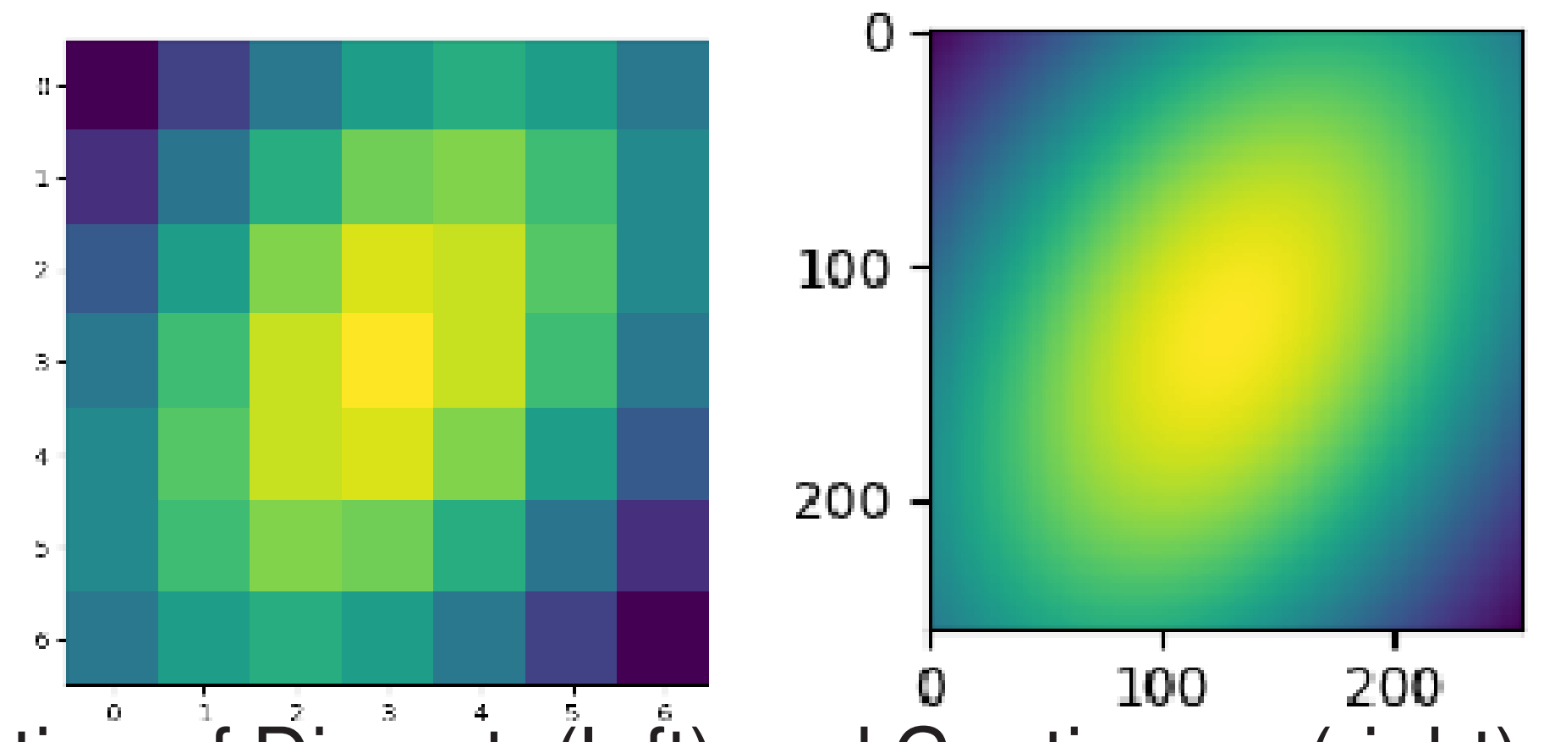


Fig. 1: Visualization of Discrete(left) and Continuous(right) Ψ. The discrete Ψ has 7 × 7 scalars, while the continuous Ψ contains infinite number of scalars and 256 × 256 are uniformly sampled and shown here.

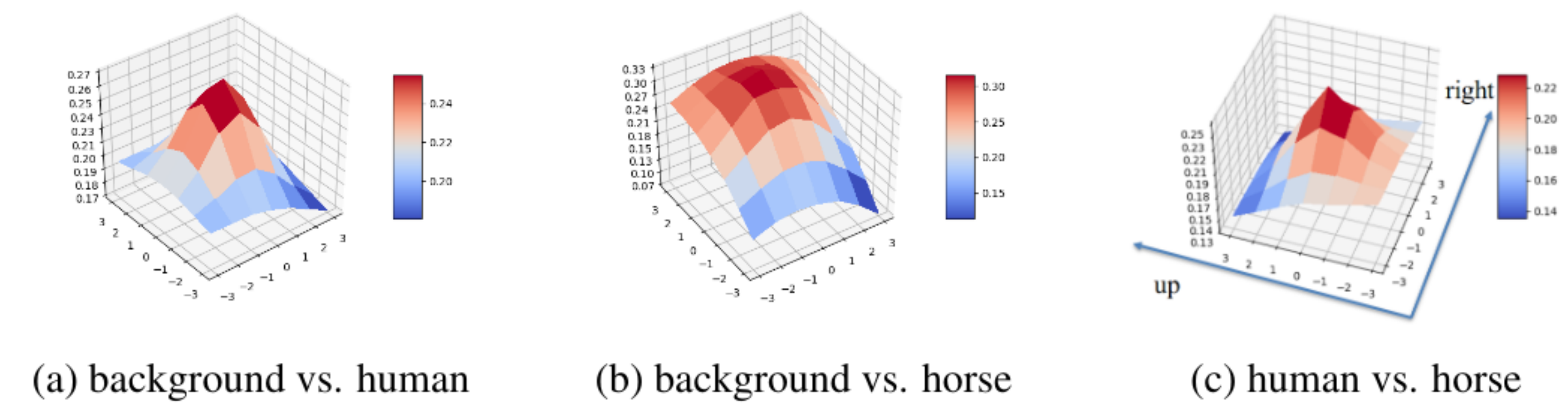
Recent progress in 3D representation learning[2] shows great capabilities of Deep Neural Network(DNN) to representing a implicit function. More specifically, for continuous function $\Psi : \mathbb{R}^2 \rightarrow \mathbb{R}$, a 7 layers MLP is used to represent the mapping from the source domain, e.g. the location of potential function, to the target domain, e.g. the potential value. Since the DNN-represented potential function is differentials, the optimization of potential function can be integrated into the end-to-end training by gradient decent. Thus, the value of potential function $\Psi(\cdot)$ in any real value location can be obtained and not limited to integer location in $[0, 6]$ anymore.

Algorithm 1: Training procedure of 7-layer MLP

```
Input  : Known potential function $\Psi$.
Output : Fitted potential function $\Psi_\theta$, where $\theta$ is parameters.
random initialization of $\theta$;
loss=1000;
while loss>0.001 do
    indexs= uniform_sample(x_min=0,x_max=1,y_min=0,y_max=1,number_sample=1000);
    /* indexs is a list of 1000 cells, and 2d index (x,y) in each cell */
    /* indexs.size=1000×2 */
    $\hat{p}$=$\Psi_\theta$(indexs);
    $p$=$\Psi$(indexs);
    loss=l2($\hat{p}$-$p$);
    $\delta_\theta$=gradient(loss,$\theta$)·learning_rate;
    $\theta$=$\theta$+$\delta_\theta$;
```

Experiments

- Discrete Experiment:
  Select images mainly includes human and horses in Microsoft COCO dataset. Kernels that encodes prior knowledge across different categories are shown. The kernel as shown in fig.2c will assign lower potential for a pixel labeled as human if it appears above a pixel of horse.



(a) background vs. human (b) background vs. horse (c) human vs. horse

Fig. 2: Visualization of spatial kernel with non-Gaussian Shape

- Continuous Experiment:
  To compare the fitting ability of discrete and continuous representation, we select 2 representative potential function, the Gaussian function and complex function defined by high resolution natural image. The Gaussian function is symmetric, smooth, and convex function, of which structure is easy to fit. On contrast, the natural image is asymmetric, zigzag, and non-convex function, which has detail local structure and is hard to fit.As it’s shown in Fig. 3, Discrete representation loses details in potential function even in most easy Gaussian case, while Continuous representation preserve lots of local detail structures in potential function. And the learned potential function can be used to decide whether it can be approximated by Gaussian or another known parametric random field.

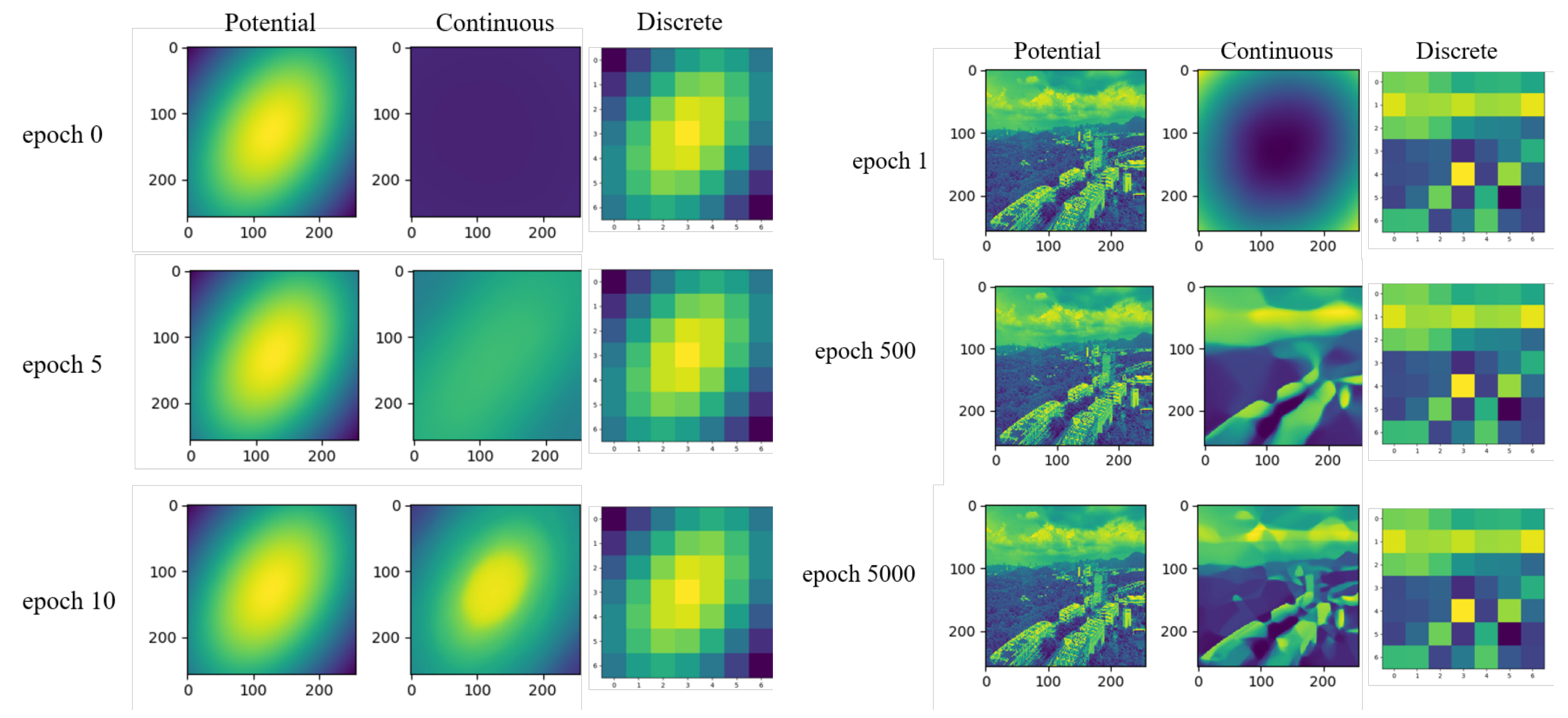


Fig. 3: left: fitting results of Gaussian function;right: fitting results of Natural Image function. Both of them shows that discrete representation loses large amount of detail structure.

References

[1] Måns Larsson, Anurag Arnab, Fredrik Kahl, Shuai Zheng, and Philip Torr. A projected gradient descent method for crf inference allowing end-to-end training of arbitrary pairwise potentials. In *International Workshop on Energy Minimization Methods in CVPR*, pages 564–579. Springer, 2017.

[2] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *arXiv preprint arXiv:2003.08934*, 2020.