Modeling Extent-of-Texture Information for Ground Terrain Recognition

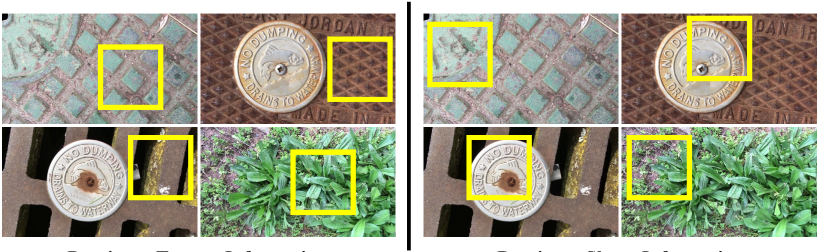
Shuvozit Ghose[a], Pinaki Nath Chowdhury[b],
Partha Pratim Roy[c], Umapada Pal[b]

[a]Institute of Engineering and Management, Kolkata, India.
[b]Indian Statistical Institute, Kolkata, India.
[c]Indian Institute of Technology, Roorkee, India.

Presenter: Pinaki Nath Chowdhury
Venue: International Conference on Pattern Recognition, 2020

Contents:

Challenges in Ground Terrain Recognition



Dominant Texture Information **Dominant Shape Information**

Ground Terrain Recognition is a difficult task due to various reasons.

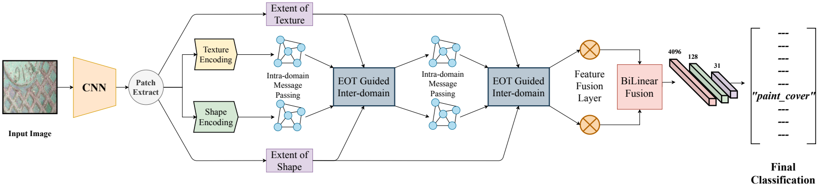
The context information varies significantly over the regions of a ground terrain image. Like some local regions posses significant texture information, while shape information is more dominant at some other parts.

Motivation

As most real-world ground terrain images show wide variations in texture and shape information at different local regions in an image, thus the classification of such realistic ground terrain images requires a more local level modeling of texture and shape information.

An Overview of our Solution

We propose a novel approach towards ground-terrain recognition via modeling the Extent-of-Texture information to establish a balance between the order-less texture and ordered-spatial information locally.



Modeling Extent-of-Texture (EoT) Information

Given an image $\mathsf{I} \in \mathbb{R}^{H \times W \times 3}$, a backbone CNN feature extractor network $G(\cdot)$ takes I and outputs latent feature representation Z. Thus,

$$\mathsf{Z} = G(\mathsf{I}; \theta_G) \tag{1}$$

Patch-extraction is performed on $\mathsf{Z} \in \mathbb{R}^{8 \times 8 \times 512}$ using a sliding window mechanism where the window size and stride is chosen as (3×3) and 1 respectively. The patch-extraction operation generates $\psi = \{\psi_i\}_{i=1}^{i=k}$, where $\psi_{\mathrm{i}} \in \mathbb{R}^{3 \times 3 \times 512}$ and k is the number of patches. Average pooling of ψ gives $\psi^* = \{\psi_i^*\}_{i=1}^{i=k}$.

Modeling Extent-of-Texture (EoT) Information (continued...)

Let $\mathrm{X} = \{x_1, x_2, x_3 ....x_k,\}$, where x_i denotes the central region of the ψ_i patch e.g. $x_i = \psi_i[2; 2; :]$ and $x_i \in \mathbb{R}^{1\times1\times512}$.

The cosine similarity between ψ^* and X describes the order-less texture information $\mathcal{T}$, where $\mathcal{T} = \{\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3 ....\mathcal{T}_k\}$ and $\mathcal{T}_i$ denotes the order-less texture information of the i^{th} patch. Therefore,

$$\psi_i^* = AvgPool(\psi_i, 3) \tag{2}$$

$$\mathcal{T}_i = \frac{\psi_i^* \cdot x_i}{||\psi_i^*||_2 \; ||x_i||_2} \tag{3}$$

$$\mathcal{T}_i = \frac{\mathcal{T}_i - \mathcal{T}_{min}}{\mathcal{T}_{max} - \mathcal{T}_{min}} \tag{4}$$

Modeling Extent-of-Texture (EoT) Information (continued...)

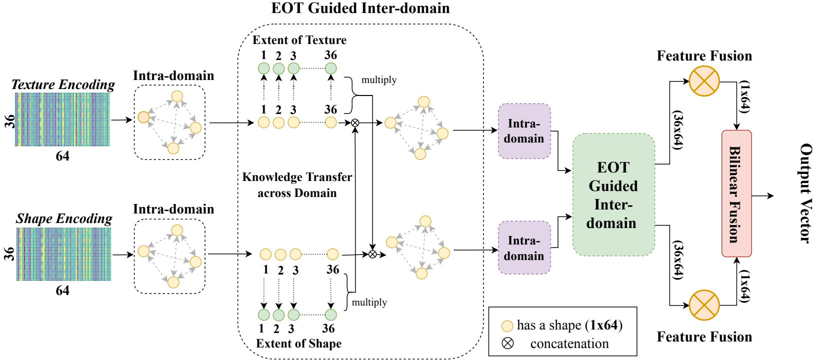
A high value of $\mathcal{T}$ indicates the presence of greater extent of the order-less texture information , whereas a small value of $\mathcal{T}$ represents higher shape information.

The ordered shape information $\mathcal{S}$, where $\mathcal{S} = \{\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3 .... \mathcal{S}_k\}$ and $\mathcal{S}_i$ denotes the ordered-spatial information of the i^{th} patch. Then,

$$\mathcal{S}_i = 1 - \mathcal{T}_i \quad (5)$$

EoT Guided Inter-domain Message Passing

The EoT Guided Inter-domain Message Passing module is used for sharing knowledge between texture and shape features to balance out the order-less texture information with ordered-spatial information.



Results

Table: Comparison of **Deep-TEN**, baseline **B1, B2, B3** and **B4** with the proposed methodology for single scale and multi scale training on GTOS-mobile [1] dataset using a pre-trained ResNet-18 module as the convolutional layer. Baseline B1 is similar to Deep Encoding Pooling Network (DEP) by Xue [1].

	Deep-TEN [2]	**B1 [1]**	**B2**	**B3**	**B4**	**Proposed Method**
Single Scale	74.22	76.07	77.81	78.55	78.93	**80.39**
Multi Scale	76.12	82.18	83.78	84.31	84.36	**85.71**

Results (continued...)

Table: Comparing Our method with several state-of-the-art methods on Describable Textures Dataset (DTD) and Materials in Context Database (MINC)

Method	DTD [3]	MINC-2500 [4]
FV-CNN [5]	72.3	63.1
Deep-TEN [2]	69.6	80.4
DEP [1]	73.2	82.0
Proposed Method	**75.7**	**85.3**

Conclusion

- we have proposed a novel approach towards ground-terrain recognition via modeling the extent of texture information to establish a balance between the order-less texture component and ordered-spatial information locally.
- The driving idea of our architecture is the modeling of context information locally.
- The proposed framework is simple and easy to implement.
- We demonstrate the effectiveness of our system by conducting experiments on publicly available ground terrain datasets.

References

- J. Xue, H. Zhang, and K. Dana, "Deep texture manifold for ground terrain recognition," in *CVPR*, 2018.
- H. Zhang, J. Xue, and K. Dana, "Deep ten: Texture encoding network," *arXiv preprint arXiv:1612.02844*, 2016.
- M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed, and A. Vedaldi, "Describing textures in the wild," in *CVPR*, 2014.
- S. Bell, P. Upchurch, N. Snavely, and K. Bala, "Material recognition in the wild with the materials in context database," in *CVPR*, 2015.
- M. Cimpoi, S. Maji, and A. Vedaldi, "Deep filter banks for texture recognition and segmentation," in *CVPR*, 2015.

Thank you

Questions?

Source Code is available at:

github.com/ShuvozitGhose/Ground-Terrain-EoT