Unifying the Visual Perception of Humans and Machines on Fine-Grained Texture Similarity

Weibo Wang, Xinghui Dong *

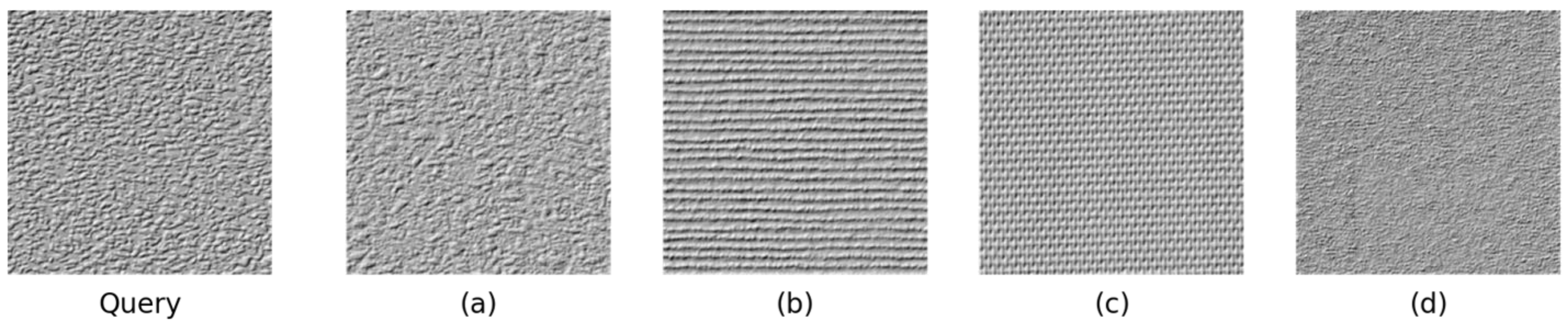
School of Computer Science and Technology, Ocean University of China

* Corresponding author

Abstract

Texture similarity plays an important role in texture analysis and material recognition. However, the prediction of perceptually-consistent fine-grained texture similarity is a challenging task. It has been found that the discrepancy between the texture representation methods and the similarity metrics utilised by humans and algorithms should account for the dilemma. To address this problem, we propose a novel Perceptually Motivated Texture Similarity Prediction Network (PMTSPN), which comprises a siamese Conformer with multi-scale bilinear pooling (SC-MSBP) and a metric learning network (MLN). The SC-MSBP learns a texture representation capturing the Higher Order Statics (HOS) in different spatial scales, while the MLN learns a similarity metric from the features which encode the short-range, long-range and lateral interactions. The PMTSPN can be trained using a set of human perceptual similarity data. Our results show that the PMTSPN produces the more consistent similarity prediction with human perception, compared with its counterparts. We attribute the promising performance to both the powerful texture representation and the effective similarity metric learnt by the PMTSPN.

Background and Motivation



Only a Weak Correlation Perception of Machines and Humans

- Each retrieval texture(a, b, c and d) has the same SSIM value (0.0264) compared with the query texture.
- However, the corresponding similarity values contained in the Isomap perceptual similarity matrix manifest great variations, which are 0.85286, 0.37449, 0.36677 and 0.544 in turn.

Texture Similarity Encounters Two Challenges

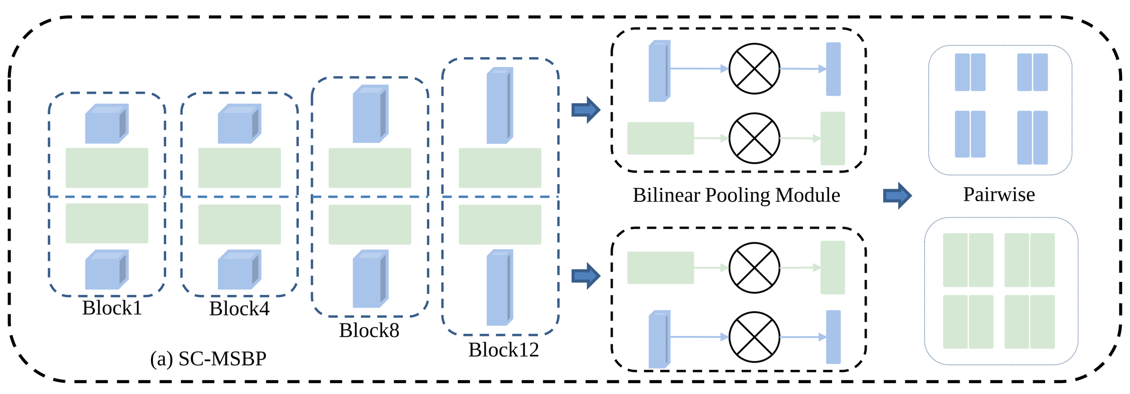
- Texture representation
- Similarity metrices
- As a result, the prediction of perceptually-consistent fine-grained texture similarity using algorithms is struggling.

Methodology

Our network comprises two subnetworks: a siamese Conformer with multi-scale bilinear pooling (SC-MSBP) and a metric learning network (MLN).

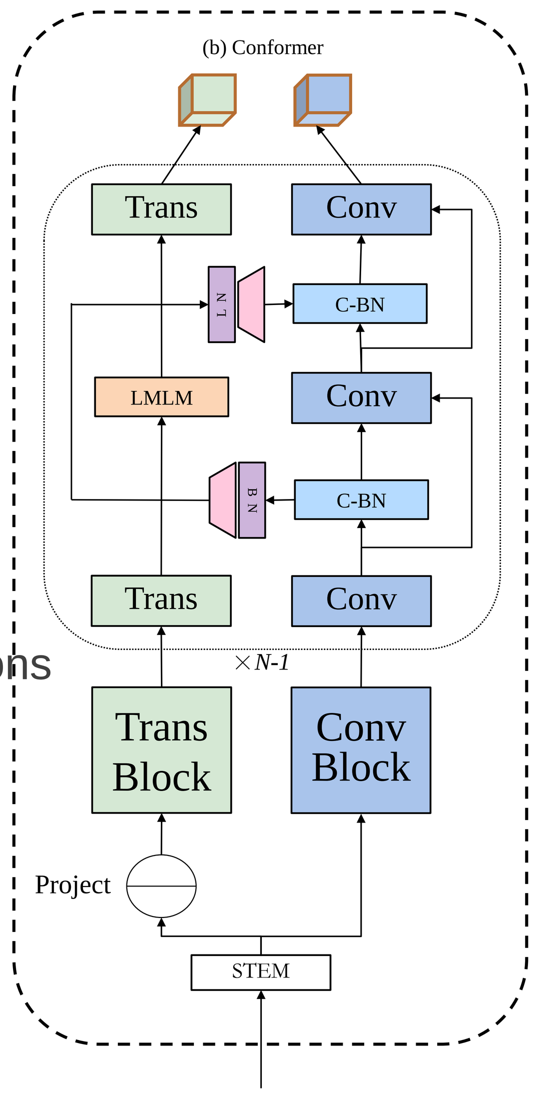
Siamese Conformer with Multi-scale Bilinear Pooling (SC-MSBP)

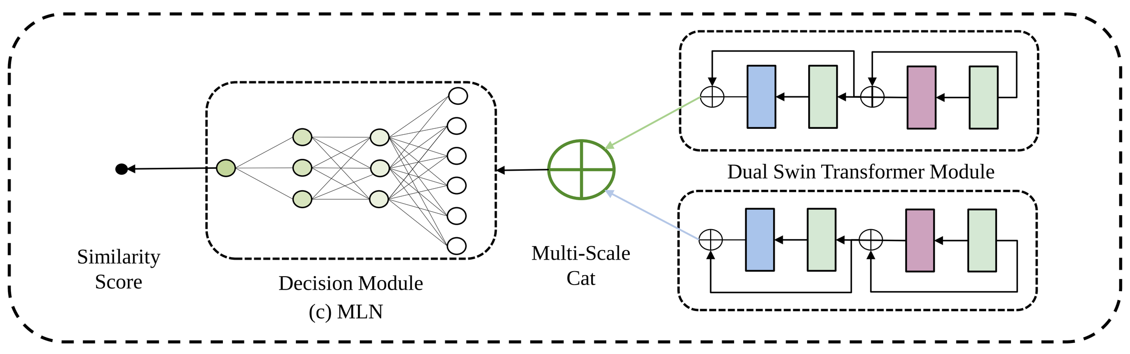
(a) Is designed for learning texture representation
(b) Is able to exploit both the local features and global representation
(c) Using bilinear pooling encoding Higher Order Statics(HOS)
(d) Using multiple resolutions simulating the Human Visual System(HVS)



Metric Learning Network (MLN)

(a) contains a dual Swin Transformer module and a decision module
(b) Compressing the features received from SC-MSBP
(c) Encode short-range, long-range and lateral interactions





Experiments

Dataset

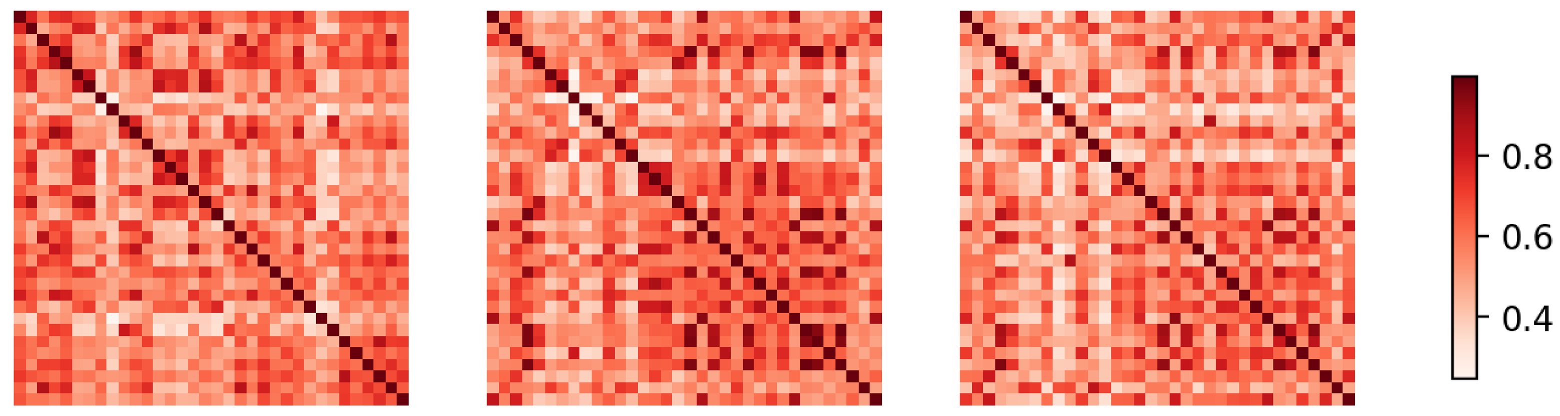
(a) Pertex Dataset
  (1) Contains 334 textures
  (2) Captured from the decorative materials
  (3) Isomap dimensionality reduction method
(b) PTD Dataset
  (1) Contains 450 textures
  (2) Generated using 23 procedural texture models

Performance Measures

- Mean Square Error (MSE)
- Pearson's Correlation Coefficient (PCC)
- Kendall's Rank Correlation Coefficient (KRCC)
- Spearman's Rank Correlation Coefficient (SRCC)

Experimental Results

Similarity Matrices



(a) Others (b)Ground-Truth (c)Ours

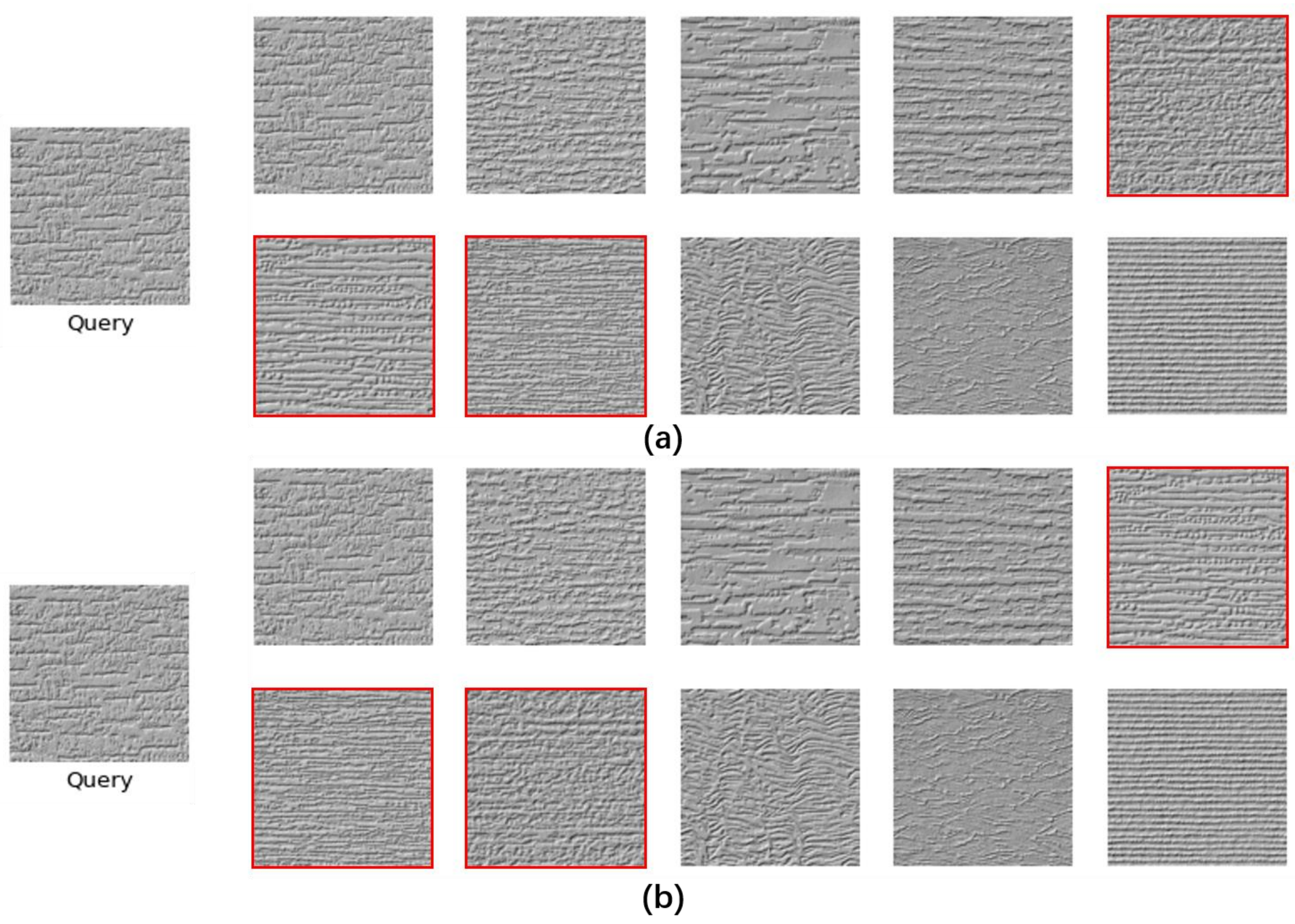
Results Obtained Using Pertex and PTD

Method		Pertex MSE	Pertex PCC	Pertex SRCC	Pertex KLCC	PTD MSE	PTD PCC	PTD SRCC	PTD KLCC
Random Forests	LBP	0.025	0.5430	-	-	0.012	0.8272	-	-
	Gabor	0.021	0.6890	-	-	0.010	0.8657	-	-
	PCANet-48D	0.026	0.6259	-	-	0.016	0.8044	-	-
	CNN-48D	0.024	0.6171	-	-	0.017	0.8048	-	-
Auto Encoder	LBP	0.030	0.3564	-	-	0.022	0.6113	-	-
	Gabor	0.019	0.6696	-	-	0.012	0.8077	-	-
	PCANet-48D	0.032	0.4687	-	-	0.013	0.7915	-	-
	CNN-48D	0.019	0.7037	-	-	0.010	0.8560	-	-
PDLF-PTSNet		0.0130	0.7805	-	-	0.0040	0.9402	-	-
PDLF-PTSNet*		0.0151	0.7949	0.7188	0.5412	0.0032	0.9546	0.9388	0.7980
DISTS**		0.0100	0.8473	0.7949	0.6048	0.0049	0.9292	0.9005	0.7327
Ours		**0.0056**	**0.9171**	**0.8783**	**0.7076**	**0.0017**	**0.9763**	**0.9628**	**0.9376**

* This model was reproduced following the work of Gao *et al.* [15]. The results shown here were selected in terms of the highest PCC value. ** Due to the unpublished training source code and the limitation on the number of data sets, we were unable to reproduce the second loss of DISTS.

Texture Retrieval Experiment

- Texture retrieval experiment using the Pertex data set
- Each test texture was used as a Query texture
- The top 10 most similar textures to the query texture were retrieved in the descending order with regard to the similarity
- Two sets of ranked textures retrieved using: (a) our best model and (b) human observers
- And red boxes indicate the difference between the two ranked lists



(a)

(b)

Conclusion

In this paper, we addressed the problem with the inconsistency between the predictions of fine-grained texture similarity by humans and algorithms. To this end, we introduced a new Perceptually Motivated Texture Similarity Prediction Network (PMTSPN). This network contains a siamese Conformer with multi-scale bilinear pooling (SC-MSBP) and a metric learning network (MLN). The two subnetworks were used to learn a texture representation which captured the Higher Order Statics (HOS) in different spatial extents, and a similarity metric from the features which encoded the short-range, long-range and lateral interactions, respectively. To our knowledge, they have not been explored for texture similarity tasks. Experimental results showed that the PMTSPN outperformed its counterparts with large margins. This promising performance should be attributed to the capability that our method learns both the powerful texture representation and the effective similarity metric.

Acknowledgements

This study was in part supported by the National Natural Science Foundation of China (NSFC) (No. 42176196) and was in part supported by the Young Taishan Scholars Program (No. tsqn201909060).