12. Appendix

12.1. Implementation

The implementation of this project is available as a Github repository at https://github.com/RosenholtzLab/StatNetExperiments.

12.2. Contrastive Learning

Our aim is to develop a function $f_\theta(x) : \mathbb{R}^F \to \mathbb{R}^D$ that pushes encoded crops from the same classes in $\mathbb{R}^F$ closer together in $\mathbb{R}^D$. On the other hand, crops in $\mathbb{R}^F$ originating from different textures are pushed further apart in $\mathbb{R}^D$.

The function f_θ is parameterized by θ. In this work θ represents the collective set of weights and biases of the neural network that are learned and adjusted during training to achieve the desired embeddings in the 50-dimensional space. For encoded textures x, the loss function employed is given through:

$$\mathrm{L}(\theta; X) = \sum_{i=1}^{P} \sum_{a=1}^{K} \left[\log \left(\sum_{\substack{p=1 \\ p \neq a}}^{K} e^{D(f_\theta(x_a^i), f_\theta(x_p^i))} \right) + \log \left(\sum_{j=1}^{P} {}_{j \neq i} \sum_{n=1}^{K} e^{m - D(f_\theta(x_a^i), f_\theta(x_n^j))} \right) \right]_+$$

Here, the first term in the bracket are all positive pairs and the last term all negatives. The two summations indicate that we consider all pairs at once. As in (Hermans et al., 2017), the distance measure used is the Euclidean distance:

$$D(f\theta(x_i), f\theta(x_j)) = \|f\theta(x_i) - f\theta(x_j)\|_2$$

12.3. Data augmentation

For our self-supervised learning, we apply several transformations to the images. We use a random vertical flip with a 0.5 probability and a horizontal flip with the same probability. At the final step, we get five crops from the adjusted image: one from each corner and one from the center. These five crops all represent one class in the dataset and the contrastive learning setup. We avoided most transformations such as blurring or jittering because they could change the statistic values. After augmenting, we encode the five cropped images using the 150-statistic set. To keep the data consistent, we normalize the statistics with the Scikit standard scaler. This helps ensure our network is not influenced by varying statistic sizes. These normalized statistics are then processed through a single-layer network with input size 150 and output size of 50.

12.4. Labeling of statistics

The labeling of statistics is systematic, driven by their statistic group and the filter of the steerable pyramid they are derived from. We follow three distinct patterns of labeling.

- Non-correlation statistics: These are indicated in the format "statistic level orientation". For instance, "end stop 1 1" refers to the end stop statistic for the first orientation at the first pyramid level.

- Correlations between neighboring scales: This follows the format "statistic (level_1, level_2) orientation", i.e. "magnitude_correlation (2,3) 3", signifying a correlation between the second and third levels for the third orientation.
- Correlations within a level across different orientations: These are denoted as "statistic level (orientation_1, orientation_2)". This structure labels the correlation occurring within a specific pyramid level but across various orientations such as magnitude correlation 1 (1,3).

12.5. Correlations in Statistics

We expected that many of the statistics measured in our analysis were likely to be correlated due to the regularities present in natural images. To investigate the degree to which correlations between different statistics are present in our analysis, we calculated the correlation between each statistic over the dataset, then used Spearman Correlation to group the statistics.

We find that indeed, many statistics are highly correlated with each other. Marginals show strong correlation with other marginals, but little correlation with other statistics. The entire population of end-stopped and magnitude statistics together have strong correlation. In addition, there are strong repeated patterns of correlation and anti-correlation between phase and magnitude statistics. Statistics of the same type and scale/level share these patterns and cluster together.

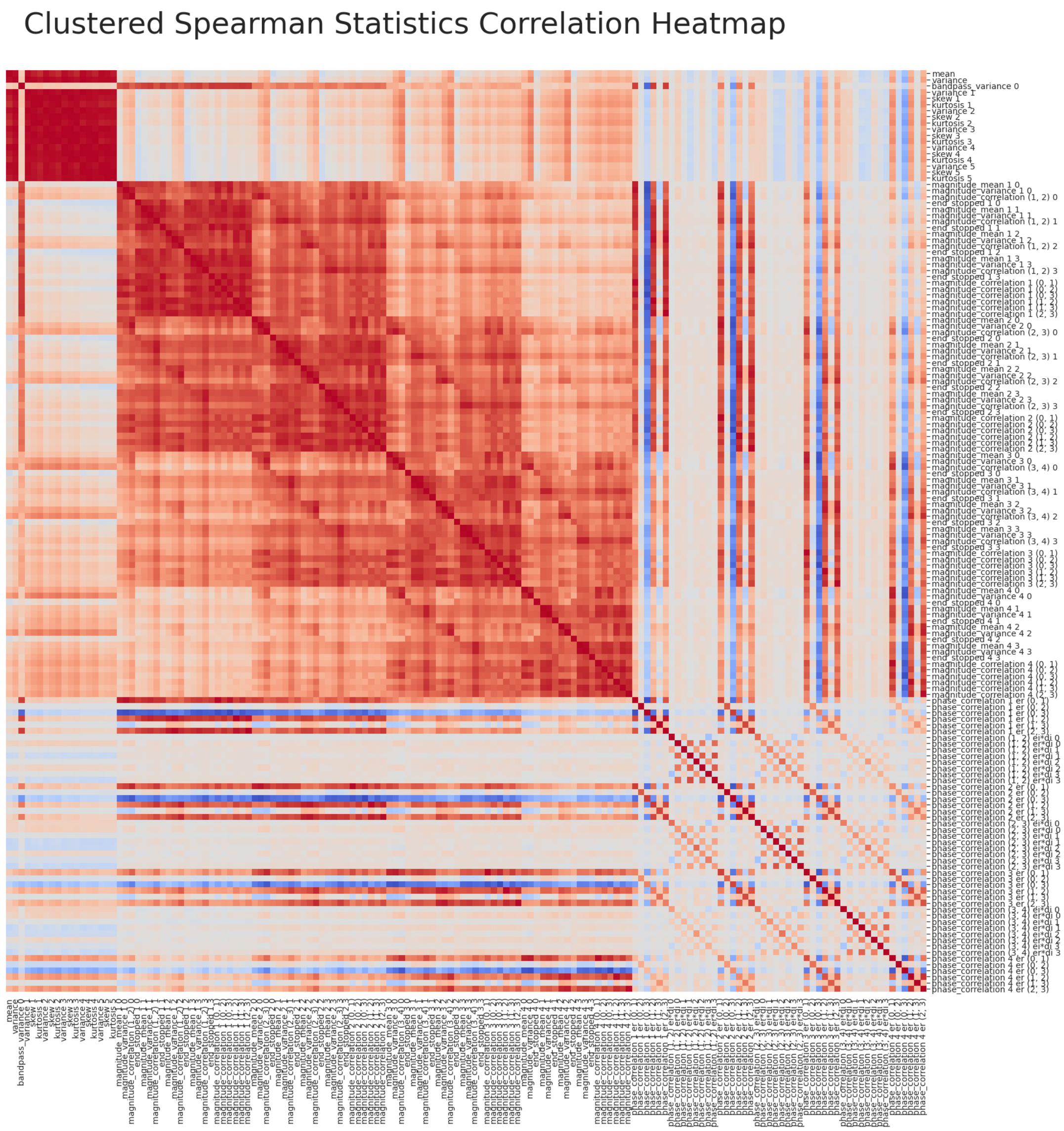

Figure 7: Correlation heatmap for all 150 statistics. Strong red color indicates positive correlation (1.0), while dark blue color anti-correlation (-1.0). There are high correlations between many statistics, especially within-group. There is also a subset of statistics that are anti-correlated or non-correlated.

12.6. Statistic Importance by Group

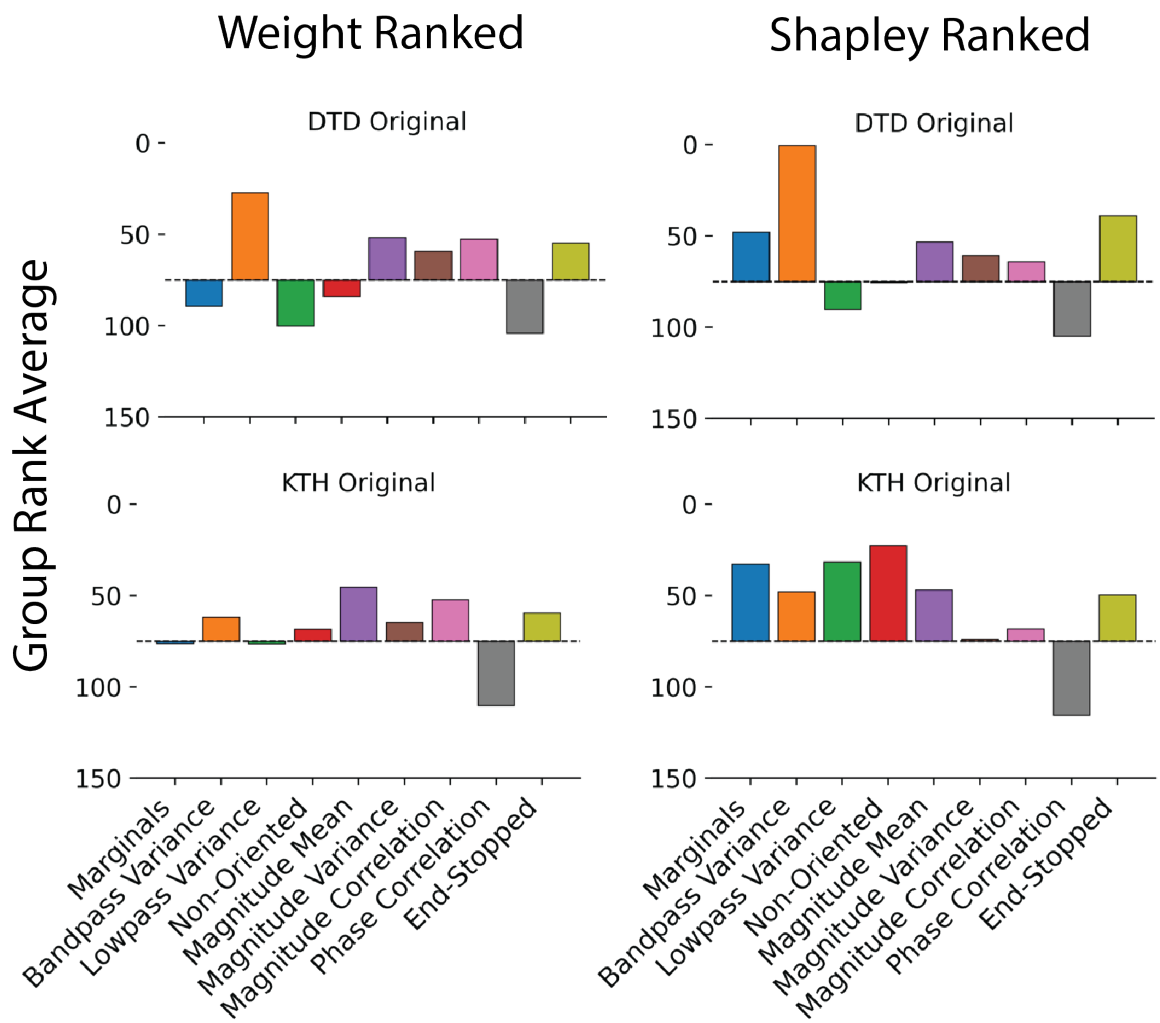

Figure 8: Mean ranking for the statistics groups with rankings based on weight (left), and Shapley values (right), for original textures of both datasets tested. Bandpass variance statistics generally rank with high importance, and phase-correlation statistics consistently rank with low importance.

12.7. Most and Least Important Statistics

Table 2: 10 most important statistics for the DTD & KTH dataset averaged over 10 seeds based on Shapley feature selection methods.

DTD		**KTH**	
Stat	Avg Rank	Stat	Avg Rank
end_stop 1 1	7.90	end_stop 3 2	5.80
end_stop 1 3	9.90	magnitude_mean 3 3	13.70
magnitude_correlation 1 (0, 2)	10.50	end_stop 3 0	14.30
end_stop 1 0	11.40	magnitude_mean 3 1	15.80
end_stop 1 2	11.90	magnitude_correlation (3, 4) 3	16.50
magnitude_correlation 1 (1, 3)	12.80	magnitude_correlation 3 (0, 2)	16.70
magnitude_mean 1 3	14.20	end_stop 2 2	17.80
magnitude_mean 1 1	14.50	magnitude_variance 3 3	19.00
magnitude_mean 1 2	19.70	magnitude_correlation (2, 3) 3	20.40
magnitude_correlation 1 (0, 3)	19.90	magnitude_mean 4 3	20.50

Table 3: 10 least important statistics for the DTD & KTH dataset averaged over 10 seeds based on Shapley feature selection methods.

DTD		**KTH**	
Stat	Avg Rank	Stat	Avg Rank
phase_correlation (2, 3) er*di 1	132.30	phase_correlation (2, 3) ei*di 1	135.90
phase_correlation (2, 3) ei*di 2	133.10	phase_correlation (1, 2) ei*di 1	136.10
phase_correlation (2, 3) ei*di 0	133.50	phase_correlation (1, 2) ei*di 3	136.10
phase_correlation (2, 3) ei*di 1	133.50	phase_correlation (3, 4) er*di 0	137.10
phase_correlation 1 er (0, 2)	134.30	end_stop 4 2	137.90
phase_correlation (2, 3) ei*di 3	135.10	phase_correlation 1 er (0, 2)	138.10
phase_correlation 2 er (0, 2)	136.20	phase_correlation 2 er (0, 2)	139.40
end_stop 4 2	136.40	phase_correlation (2, 3) ei*di 0	139.80
phase_correlation (3, 4) ei*di 0	138.20	phase_correlation (2, 3) er*di 0	140.00
phase_correlation (3, 4) er*di 0	139.90	phase_correlation (2, 3) er*di 2	140.60