# Image Quality Assessment by Comparing CNN Features Between Images

**Seyed Ali Amirshahi[*†], Marius Pedersen[†], Stella X. Yu[*]**

January 30th, 2017

[*] UC Berkeley / International Computer Science Institute (ICSI), Berkeley, USA

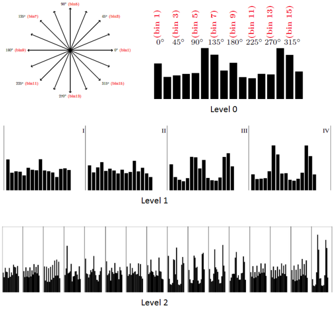
[†] Norwegian University of Science and Technology, Gjøvik, Norway

# Introduction

- Over the years a high number of different image quality methods have been introduced.
- When it comes to performing well across databases and distortions there exist room for improvement.
- Most image quality metrics use a limited number of handcrafted features [Amirshahi and Larabi, 2011, Pedersen and Hardeberg, 2009].
- Taking advantage of CNNs, our approach not only takes low level features into account but it also compares mid and high level features providing a more precise and accurate metric.

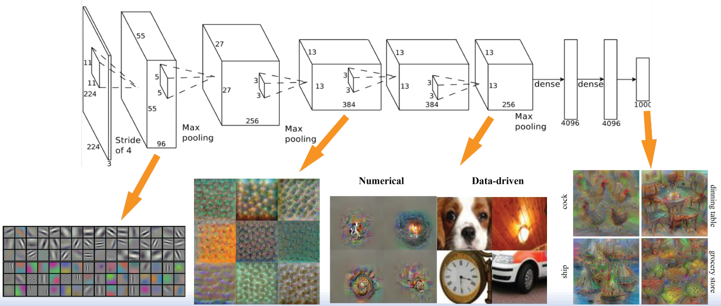
# Self-similarity in images

- Recently, different measures of self-similarity were introduced in the field of computational aesthetics [Amirshahi, 2015].
- The mentioned methods take a pyramidal approach in which HOGs [Dalal and Triggs, 2005] in different regions are compared to smaller sub-regions in the image.
- In this study, we extend this work to evaluate the similarity seen between two given images, the test and the reference image.

# Overview of the proposed metric

- In the proposed approach we use a pre-trained AlexNet [Krizhevsky et al., 2012] model on the ImageNet dataset [Deng et al., 2009].
- Feature maps extracted from the test and reference image at different convolutional layers are compared in various spatial levels.
- In other words, we use the strength of the feature maps as bin entries in the pyramidal approach to evaluate the similarity between two given images.




Conv 1: Edge+Blob Conv 3: Texture Conv 5: Object Parts Fc8: Object Classes

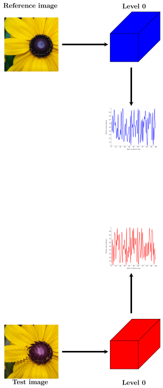
# Proposed image quality metric

- The following steps are taken in the calculation of the proposed image quality metric:

  1. For the test image $\mathcal{I}_T$, at layer $n$ and level $L$ of the spatial pyramid we calculate histogram

$$\boldsymbol{h}(\mathcal{I}_T, n, L) = (\sum_{i=1}^{X}\sum_{j=1}^{Y}\mathcal{F}(\mathcal{I}_T, n, L, 1)(i,j), \cdots,$$
$$\sum_{i=1}^{X}\sum_{j=1}^{Y}\mathcal{F}(\mathcal{I}_T, n, L, z)(i,j), \cdots, \sum_{i=1}^{X}\sum_{j=1}^{Y}\mathcal{F}(\mathcal{I}_T, n, L, M)(i,j)).$$

     - In the case of the AlexNet model, 96 for the first, 256 for the second, 384 for the third and fourth, and 256 for the fifth convolutional layer.

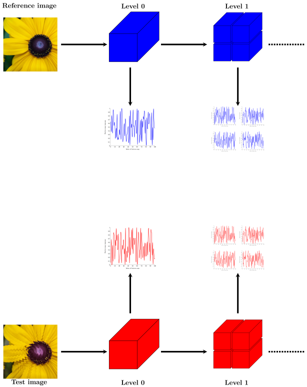


# Proposed image quality metric

- The following steps are taken in the calculation of the proposed image quality metric:

  2. To maintain the pyramidal nature, the division of the sub-regions and calculation of ***h*** will continue till the smallest side of the smallest sub-region is equal or larger than seven pixels.

     - In the AlexNet model this results in the third level for the first convolutional layer, the second level for the second layer, and the first level for the third, fourth, and fifth layers.

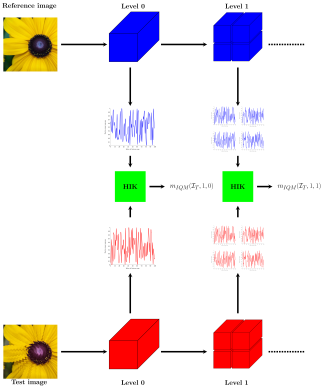


# Proposed image quality metric

- The following steps are taken in the calculation of the proposed image quality metric:

  3. Using the Histogram Intersection Kernel [Barla et al., 2002], the quality of the test image ($\mathcal{I}_T$) at level $L$ of the spatial pyramid for the $n^{\text{th}}$ convolutional layer is calculated by,

$$m_{IQM}(\mathcal{I}_T, n, L) = d_{HIK}(\boldsymbol{h}(\mathcal{I}_T, n, L)), \boldsymbol{h}(\mathcal{I}_R, n, L))) = \sum_{i=1}^{n} \min(h_i(\mathcal{I}_T, n, L)), h_i(\mathcal{I}_R, n, L)).$$



## Proposed image quality metric

- The following steps are taken in the calculation of the proposed image quality metric:

  4. For each convolutional layer $n$ in the test image, we introduce the quality vector

  $$\boldsymbol{m}_{IQM}(\mathcal{I}_T, n) = (m_{IQM}(\mathcal{I}_T, n, 1), m_{IQM}(\mathcal{I}_T, n, 2), \cdots, m_{IQM}(\mathcal{I}_T, n, z), \cdots, m_{IQM}(\mathcal{I}_T, n, L)),$$

  which is the result of the concatenation of $m_{IQM}(\mathcal{I}_T, n, l)$ values for all the levels in the spatial pyramid.

# Proposed image quality metric

- The following steps are taken in the calculation of the proposed image quality metric:
  5. The quality of the test image $\mathcal{I}_T$ at the $n^{\text{th}}$ convolutional layer is calculated by

$$IQ(\mathcal{I}_T, n) = \frac{1 - \sigma(\boldsymbol{m}_{IQM}(\mathcal{I}_T, n))}{\sum_{l=1}^{L} \frac{1}{l}} \sum_{l=1}^{L} \frac{1}{l} \cdot m_{IQM}(\mathcal{I}_T, n, l).$$

# Proposed image quality metric

- The following steps are taken in the calculation of the proposed image quality metric:
  5. The quality of the test image $\mathcal{I}_T$ at the $n^{\text{th}}$ convolutional layer is calculated by

$$IQ(\mathcal{I}_T, n) = \frac{1 - \sigma(\boldsymbol{m}_{IQM}(\mathcal{I}_T, n))}{\sum_{l=1}^{L} \frac{1}{l}} \sum_{l=1}^{L} \frac{1}{l} \cdot m_{IQM}(\mathcal{I}_T, n, l).$$

# Proposed image quality metric

- The following steps are taken in the calculation of the proposed image quality metric:
  5. The quality of the test image $\mathcal{I}_T$ at the $n^{\text{th}}$ convolutional layer is calculated by

$$IQ(\mathcal{I}_T, n) = \frac{1 - \sigma(\boldsymbol{m}_{IQM}(\mathcal{I}_T, n))}{\sum_{l=1}^{L} \frac{1}{l}} \sum_{l=1}^{L} \frac{1}{l} \cdot m_{IQM}(\mathcal{I}_T, n, l).$$

# Proposed image quality metric

- The following steps are taken in the calculation of the proposed image quality metric:
  5. The quality of the test image $\mathcal{I}_T$ at the $n^{\text{th}}$ convolutional layer is calculated by

$$IQ(\mathcal{I}_T, n) = \frac{1 - \sigma(\boldsymbol{m}_{IQM}(\mathcal{I}_T, n))}{\sum_{l=1}^{L} \frac{1}{l}} \sum_{l=1}^{L} \frac{1}{l} \cdot m_{IQM}(\mathcal{I}_T, n, l).$$

# Proposed image quality metric

- The following steps are taken in the calculation of the proposed image quality metric:

  5. The quality of the test image $\mathcal{I}_T$ at the $n^{\text{th}}$ convolutional layer is calculated by

$$IQ(\mathcal{I}_T, n) = \frac{1 - \sigma(\boldsymbol{m}_{IQM}(\mathcal{I}_T, n))}{\sum_{l=1}^{L} \frac{1}{l}} \sum_{l=1}^{L} \frac{1}{l} \cdot m_{IQM}(\mathcal{I}_T, n, l).$$

# Proposed image quality metric

- The following steps are taken in the calculation of the proposed image quality metric:

6. To link the quality values calculated at different convolutional layers, we use the geometric mean

$$IQ(\mathcal{I}_T) = \prod_{n=1}^{5} IQ(\mathcal{I}_T, n).$$

## Datasets used

- In our experiments we used the following datasets:
  - Colourlab Image Database: Image Quality (CID:IQ) [Liu et al., 2014].
  - LIVE Image Quality Assessment Database release 2 (LIVE2) [Sheikh et al., 2006, Sheikh et al., 2005].
  - Computational and Subjective Image Quality (CSIQ) [Larson and Chandler, 2010].
  - Tampere Image Database (TID2013) [Ponomarenko et al., 2015].

## Datasets used

- In our experiments we used the following datasets:

| | # reference image | # test image | # distortions | # observers |
|---|---|---|---|---|
| CID:IQ | 23 | 690 | 6 | 17 |
| CSIQ | 30 | 866 | 6 | 35 |
| TID2013 | 25 | 3000 | 24 | 971 |
| LIVE2 | 29 | 982 | 5 | – |

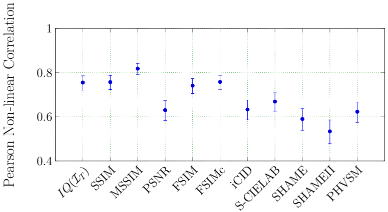
# Datasets used

- In our experiments we used the following datasets:

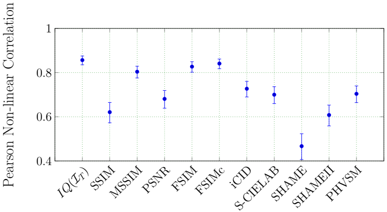
| | # reference image | # test image | # distortions | # observers | Correlation |
|---|---|---|---|---|---|
| CID:IQ | 23 | 690 | 6 | 17 | 0.76 and 0.87 |
| CSIQ | 30 | 866 | 6 | 35 | 0.92 |
| TID2013 | 25 | 3000 | 24 | 971 | 0.84 |
| LIVE2 | 29 | 982 | 5 | – | 0.91 |

# Comparing results among different metrics

- CID:IQ dataset



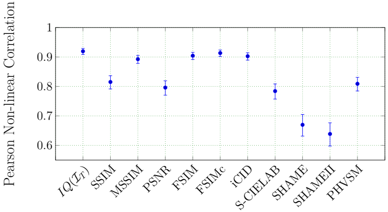

50cm viewing distance



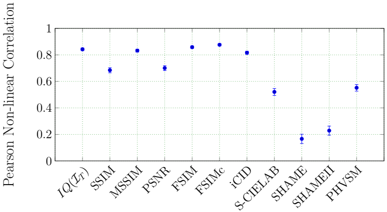

100cm viewing distance

# Comparing results among different metrics

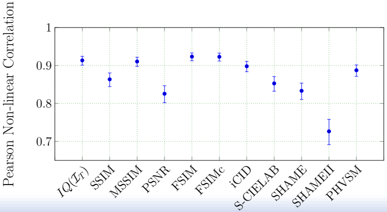
- CSIQ dataset




- TID2013 dataset




- LIVE2 dataset



# Comparing results between different distortions

- CSIQ dataset:

| | $IQ(\mathcal{I}_T)$ | SSIM | MSSIM | PSNR | FSIM | FSIMc | iCID | S-CIELAB | SHAME | SHAME II | PHVSM |
|---|---|---|---|---|---|---|---|---|---|---|---|
| All images | **0.92** | 0.82 | 0.89 | 0.80 | 0.90 | 0.91 | 0.90 | 0.78 | 0.67 | 0.64 | 0.81 |
| Gaussian blurring | **0.95** | 0.90 | 0.87 | 0.91 | 0.89 | 0.89 | 0.94 | 0.92 | 0.58 | 0.92 | 0.92 |
| Global contrast | **0.96** | 0.85 | **0.96** | 0.94 | 0.92 | 0.93 | **0.96** | 0.92 | 0.80 | 0.83 | 0.94 |
| JPEG | **0.98** | 0.947 | **0.98** | 0.89 | **0.98** | **0.98** | 0.97 | 0.97 | 0.92 | 0.91 | 0.97 |
| JPEG 2000 | **0.98** | 0.92 | **0.98** | 0.95 | **0.98** | **0.98** | 0.96 | 0.95 | 0.87 | 0.66 | **0.98** |
| Additive pink Gaussian noise | 0.94 | 0.93 | 0.95 | 0.95 | 0.93 | 0.94 | **0.96** | 0.94 | 0.62 | 0.79 | **0.96** |

In each row the highest correlation is shown by red, the second highest by blue and the third highest by green.

# Comparing results between different distortions

- LIVE2 dataset:

| | $IQ(\mathcal{I}_T)$ | SSIM | MSSIM | PSNR | FSIM | FSIMc | iCID | S-CIELAB | SHAME | SHAME II | PHVSM |
|---|---|---|---|---|---|---|---|---|---|---|---|
| All images | 0.91 | 0.86 | 0.91 | 0.83 | **0.92** | **0.92** | 0.90 | 0.85 | 0.83 | 0.73 | 0.89 |
| Blur | **0.98** | 0.87 | 0.96 | 0.78 | 0.97 | 0.97 | 0.93 | 0.83 | 0.87 | 0.91 | 0.92 |
| Fast fading rayleigh | 0.91 | **0.95** | **0.95** | 0.89 | **0.95** | **0.95** | 0.94 | 0.82 | 0.80 | 0.74 | 0.89 |
| JPEG 2000 | **0.96** | 0.94 | **0.96** | 0.90 | **0.96** | **0.96** | 0.95 | 0.90 | 0.87 | 0.81 | 0.95 |
| JPEG | 0.94 | 0.94 | 0.94 | 0.85 | **0.95** | **0.95** | 0.94 | 0.92 | 0.91 | 0.88 | 0.94 |
| White noise | **0.99** | 0.98 | 0.98 | **0.99** | 0.97 | 0.98 | 0.97 | 0.98 | 0.96 | 0.92 | **0.99** |

In each row the highest correlation is shown by red, the second highest by blue and the third highest by green.

# Comparing results between different distortions

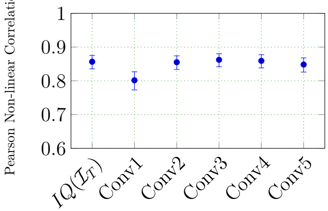
- TID2013 dataset:

| | $IQ(\mathcal{I}_T)$ | SSIM | MSSIM | PSNR | FSIM | FSIMc | iCID | S-CIELAB | SHAME | SHAME II | PHVSM |
|---|---|---|---|---|---|---|---|---|---|---|---|
| All images | 0.84 | 0.68 | 0.83 | 0.70 | 0.86 | **0.88** | 0.82 | 0.52 | 0.17 | 0.23 | 0.55 |
| Additive gaussian noise | 0.85 | 0.68 | 0.81 | 0.71 | 0.85 | **0.87** | 0.81 | 0.53 | 0.24 | 0.31 | 0.69 |
| JPEG compression | 0.88 | 0.71 | 0.85 | 0.69 | 0.88 | **0.91** | 0.79 | 0.51 | 0.67 | 0.26 | 0.69 |
| JPEG 2000 transmission errors | **0.88** | 0.66 | 0.77 | 0.67 | 0.78 | 0.81 | 0.81 | 0.60 | 0.28 | 0.55 | 0.64 |
| Mean shift (intensity shift) | **0.93** | 0.69 | 0.86 | 0.72 | 0.91 | 0.92 | 0.80 | 0.66 | 0.26 | 0.32 | 0.71 |
| Lossy compression of noisy images | **0.88** | 0.72 | 0.87 | 0.68 | 0.85 | **0.88** | 0.84 | 0.52 | 0.15 | 0.32 | 0.65 |
| Image color quantization with dither | **0.90** | 0.72 | 0.85 | 0.72 | 0.88 | **0.90** | 0.84 | 0.54 | 0.20 | 0.31 | 0.70 |

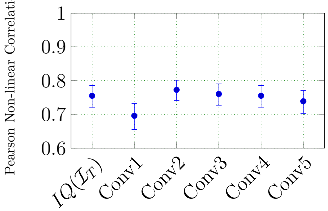
In each row the highest correlation is shown by red, the second highest by blue and the third highest by green.

# Correlation results in convolutional layers

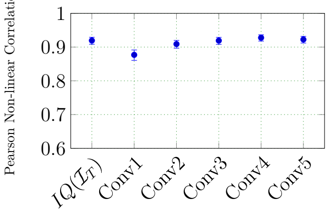
- CID:IQ in 100cm



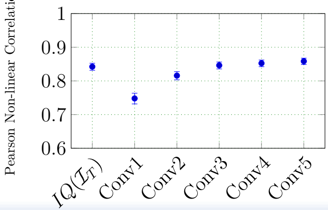

- CID:IQ in 50cm



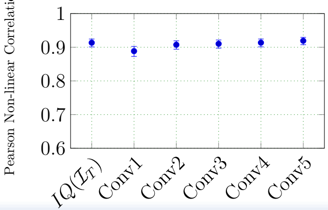

- CSIQ




- TID2013




- LIVE2



# Other CNN models

- We also evaluated the performance of our proposed metric using other CNN models, VGG 16 and VGG 19 [Simonyan and Zisserman, 2014].

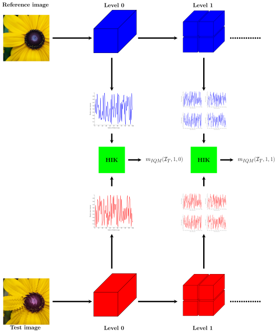
| | AlexNet | VGG 16 | VGG 19 |
|---|---|---|---|
| CID:IQ 100cm | 0.87 | 0.86 | 0.86 |
| CID:IQ 50cm | 0.76 | 0.76 | 0.77 |
| CSIQ | 0.92 | 0.94 | 0.94 |
| TID2013 | 0.84 | 0.85 | 0.85 |
| LIVE2 | 0.91 | 0.96 | 0.96 |

## Effects of convolutional layers on the performance

| | VGG 16 | | | VGG 19 | | |
|---|---|---|---|---|---|---|
| | 4 Layers | 8 Layers | 12 Layers | 5 Layers | 10 Layers | 15 Layers |
| CID:IQ 100cm | 0.80 | 0.85 | 0.86 | 0.82 | 0.85 | 0.86 |
| CID:IQ 50cm | 0.71 | 0.76 | 0.77 | 0.73 | 0.77 | 0.77 |
| CSIQ | 0.89 | 0.93 | 0.94 | 0.92 | 0.93 | 0.94 |
| TID2013 | 0.73 | 0.78 | 0.84 | 0.78 | 0.81 | 0.83 |
| LIVE2 | 0.95 | 0.95 | 0.96 | 0.95 | 0.95 | 0.96 |

# Conclusion

In this work we introduced a new full reference Image Quality Metric.




- The metric is based on comparing feature maps at different convolutional layers on different spatial levels.
- Since we are working with pre-trained networks the computational time and power needed is low.
- The proposed approach outperforms the state-of-the-art image quality metrics in most datasets and distortion types.
- We found out that the deeper the network the more accuracy we have but then we would need more time and computational power for our calculations as well.

# References I


**Amirshahi, S. A. (2015).**
*Aesthetic Quality Assessment of Paintings.*
Verlag Dr. Hut.

**Amirshahi, S. A., Koch, M., Denzler, J., and Redies, C. (2012).**
PHOG analysis of self-similarity in aesthetic images.
In *IS&T/SPIE Electronic Imaging*, pages 82911J–82911J. International Society for Optics and Photonics.

**Amirshahi, S. A. and Larabi, M. C. (2011).**
Spatial-temporal video quality metric based on an estimation of QoE.
In *Quality of Multimedia Experience (QoMEX), 2011 Third International Workshop on*, pages 84–89. IEEE.

**Amirshahi, S. A., Redies, C., and Denzler, J. (2013).**
How self-similar are artworks at different levels of spatial resolution?
In *Symposium on Computational Aesthetics*, pages 93–100. ACM.

**Barla, A., Franceschi, E., Odone, F., and Verri, A. (2002).**
Image kernels.
In *Pattern Recognition with Support Vector Machines*, pages 83–96. Springer Berlin Heidelberg.

**Bosch, A., Zisserman, A., and Munoz, X. (2007).**
Representing shape with a spatial pyramid kernel.
In *Proceedings of the 6th ACM international conference on Image and video retrieval*, pages 401–408. ACM.

# References II

**Dalal, N. and Triggs, B. (2005).**
Histograms of oriented gradients for human detection.
In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 886–893. IEEE.

**Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009).**
Imagenet: A large-scale hierarchical image database.
In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. IEEE.

**Gong, M. and Pedersen, M. (2012).**
Spatial pooling for measuring color printing quality attributes.
*Journal of Visual Communication and Image Representation*, 23(5):685—-696.

**Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012).**
Imagenet classification with deep convolutional neural networks.
In *Advances in neural information processing systems*, pages 1097–1105.

**Larson, E. C. and Chandler, D. M. (2010).**
Most apparent distortion: full-reference image quality assessment and the role of strategy.
*Journal of Electronic Imaging*, 19(1):011006–011006.

**Liu, X., Pedersen, M., and Hardeberg, J. Y. (2014).**
CID:IQ–a new image quality database.
In *International Conference on Image and Signal Processing*, pages 193–202. Springer International Publishing.

# References III


**Pedersen, M. and Hardeberg, J. Y. (2009).**
A new spatial hue angle metric for perceptual image difference.
In *International Workshop on Computational Color Imaging*, pages 81–90. Springer.

**Pedersen, M. and Hardeberg, J. Y. (2012).**
A new spatial filtering based image difference metric based on hue angle weighting.
*Journal of Imaging Science and Technology*, 56(5):50501–1.

**Ponomarenko, N., Jin, L., Ieremeiev, O., Lukin, V., Egiazarian, K., Astola, J., Vozel, B., Chehdi, K., Carli, M., Battisti, F., et al. (2015).**
Image database tid2013: Peculiarities, results and perspectives.
*Signal Processing: Image Communication*, 30:57–77.

**Redies, C., Amirshahi, S. A., Koch, M., and Denzler, J. (2012).**
PHOG-derived aesthetic measures applied to color photographs of artworks, natural scenes and objects.
In *Computer Vision–ECCV 2012. Workshops and Demonstrations*, pages 522–531. Springer.

**Sheikh, H. R., Sabir, M. F., and Bovik, A. C. (2006).**
A statistical evaluation of recent full reference image quality assessment algorithms.
*IEEE Transactions on image processing*, 15(11):3440–3451.

**Sheikh, H. R., Wang, Z., Cormack, L., and Bovik, A. C. (2005).**
Live image quality assessment database release 2.

# References IV


**Simonyan, K. and Zisserman, A. (2014).**
Very deep convolutional networks for large-scale image recognition.
*arXiv preprint arXiv:1409.1556.*

**Torkamani-Azar, F. and Amirshahi, S. A. (2007).**
A new approach for image quality assessment using svd.
In *Signal Processing and Its Applications, 2007. ISSPA 2007. 9th International Symposium on*, pages 1–4. IEEE.

**Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004).**
Image quality assessment: from error visibility to structural similarity.
*IEEE transactions on image processing*, 13(4):600–612.

# Thank You

This work was supported by a fellowship within the FITweltweit program of the German Academic Exchange Service (DAAD).