Image Quality Assessment using Synthetic Images

Pavan C Madhusudana [1] Neil Birkbeck [2] Yilin Wang [2] Balu Adsumilli [2] Alan C. Bovik [1]

[1] The University of Texas at Austin, USA

[2] Google Inc., USA

Video/Audio Quality in Computer Vision, WACV 2022

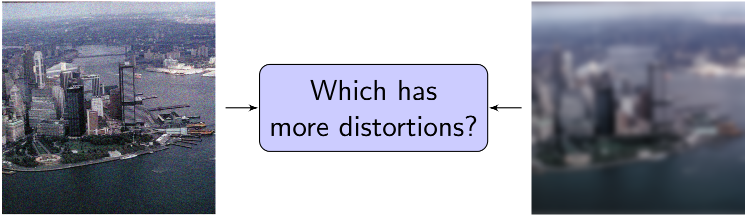
Waikoloa, Hawaii, USA

Outline of the Talk

Outline of the Talk

Problem Definition - Unsupervised IQA

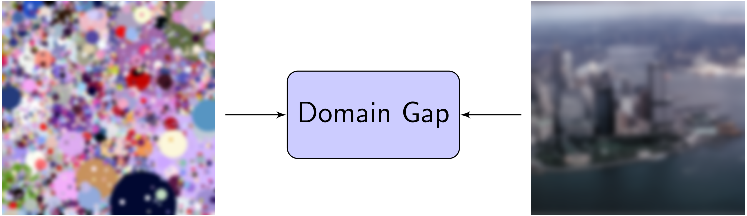
- Task - blind image quality prediction



- Learning problem - supervised learning
  - Requirement - large labeled IQA datasets
- Goal : Unsupervised feature learning for IQA
  - Distortion discrimination using contrastive learning (CONTRIQUE)
- Can synthetic images be used for training?

Motivation

- Problem setup
  - Training - synthetic data
  - Testing - real data, no additional fine-tuning on real data
- Studying the effects of synthetic data
  - Impact of domain gap between real and synthetic data
  - Significance of semantic information to quantify artifacts

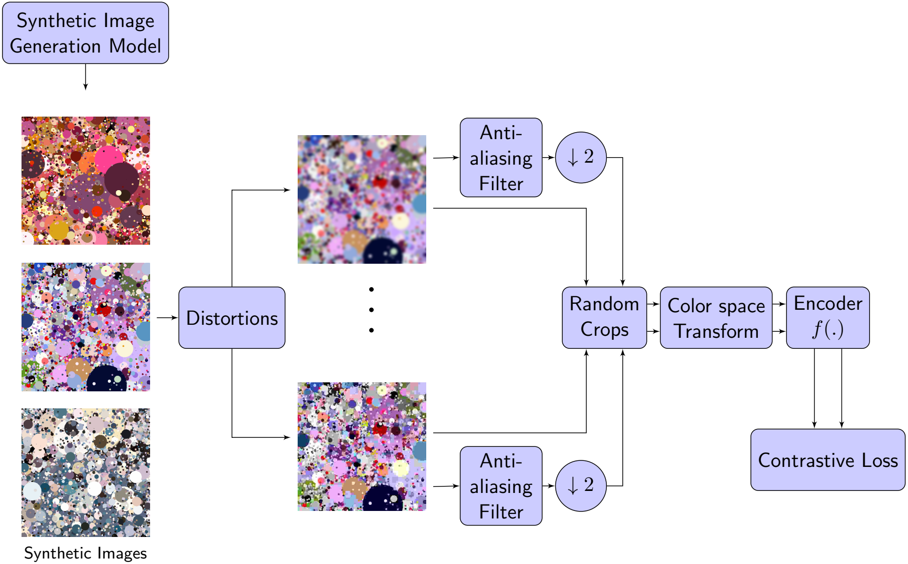


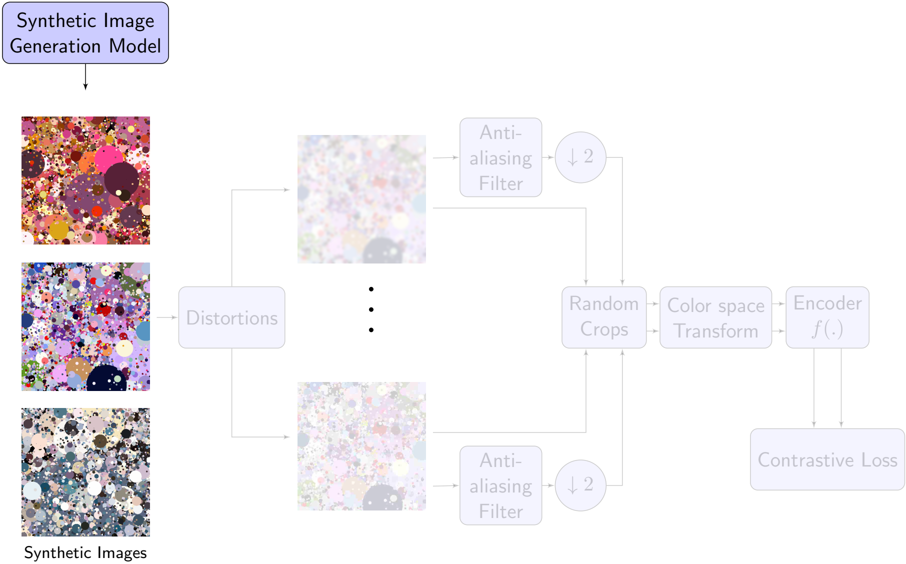
Synthetic Image

Real Image

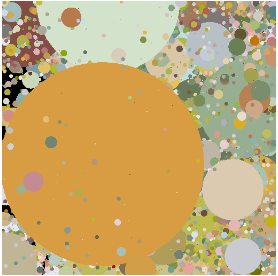
Outline of the Talk

Method Overview

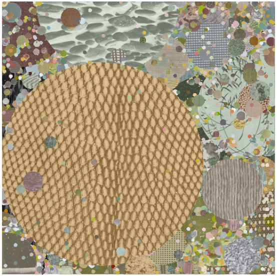



Synthetic Image Generation Model
Distortions
Anti-aliasing Filter
↓ 2
Anti-aliasing Filter
↓ 2
Random Crops
Color space Transform
Encoder f(.)
Contrastive Loss
Synthetic Images

Synthetic Datasets



DL Image

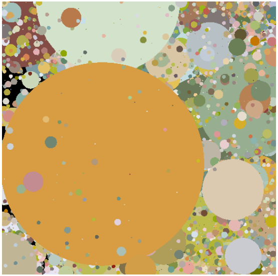
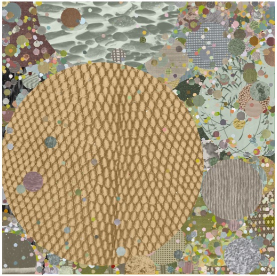


Textured DL Image

Anime Image

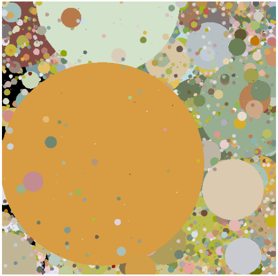
- Dead Leaves (DL) - primitive model, less semantic information
  - Image statistics - similar to natural images
  - Obtained by superposing discs of random radii and color

Synthetic Datasets - Texture

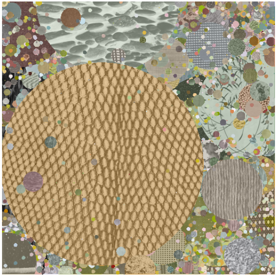
		
DL Image	Textured DL Image	Anime Image

- Textured Dead Leaves (DL) - addition of textures
  - Image statistics - more closer to natural images
  - Texture addition - improves model performance

Synthetic Datasets - Animation Images



DL Image

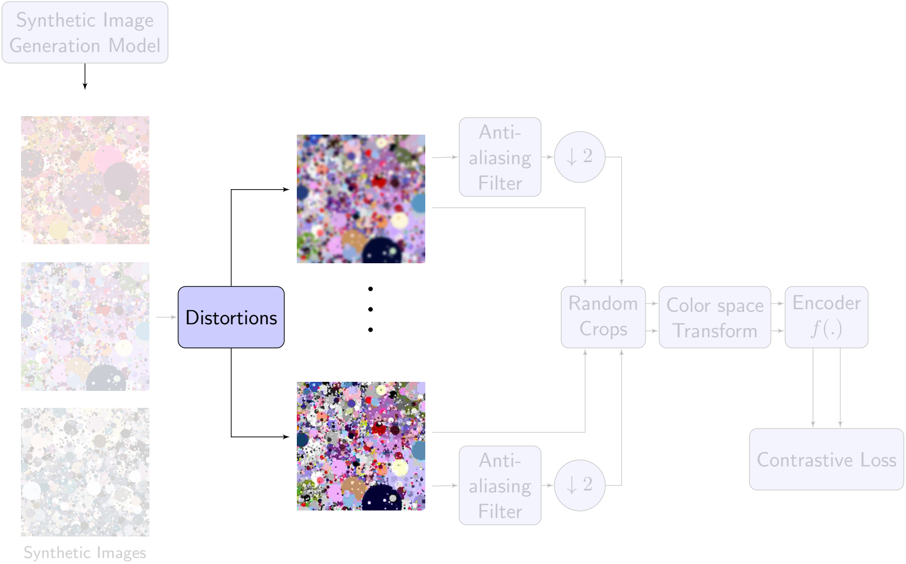


Textured DL Image

Anime Image

- Anime images
    - Contain more semantic information than DL images
    - Image generation model - simple to sophisticated methods
        - Our experiments - Danbooru dataset

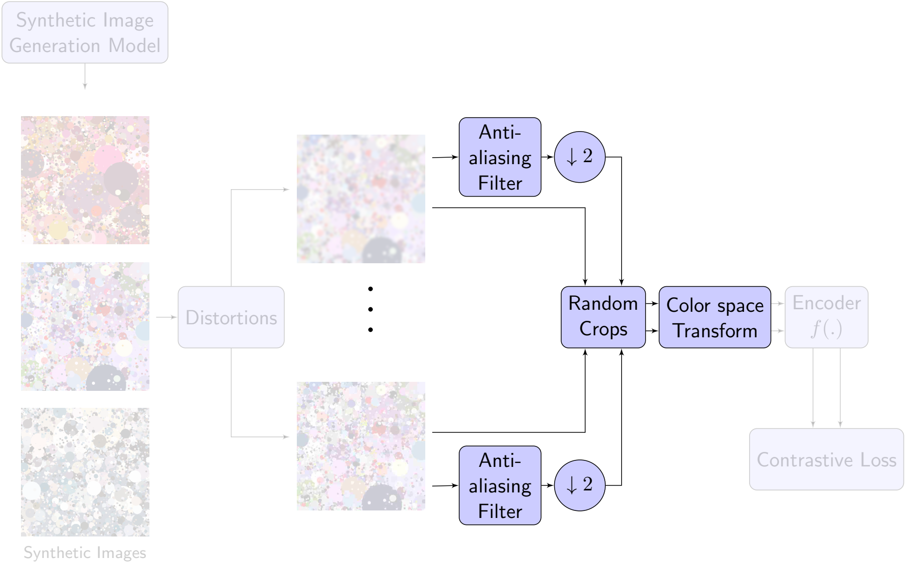
Auxiliary Task



Auxiliary Task

- Self-supervised learning - training on unlabeled data
- Auxiliary task - closely related to original task
  - Labels - easily available/generated
- Distortion type and degree discrimination
  - Synthetic image s be distorted by $d^i, i \in \{1, \ldots, D\}$ with degradation degree $l^{ij}, j \in \{1, \ldots, L^i\}$ resulting in a distorted image $\tilde{s}_i^j$.
- Objective - determine d^i and l^{ij} from given corrupted image $\tilde{s}_i^j$
- Training - multi-class classification

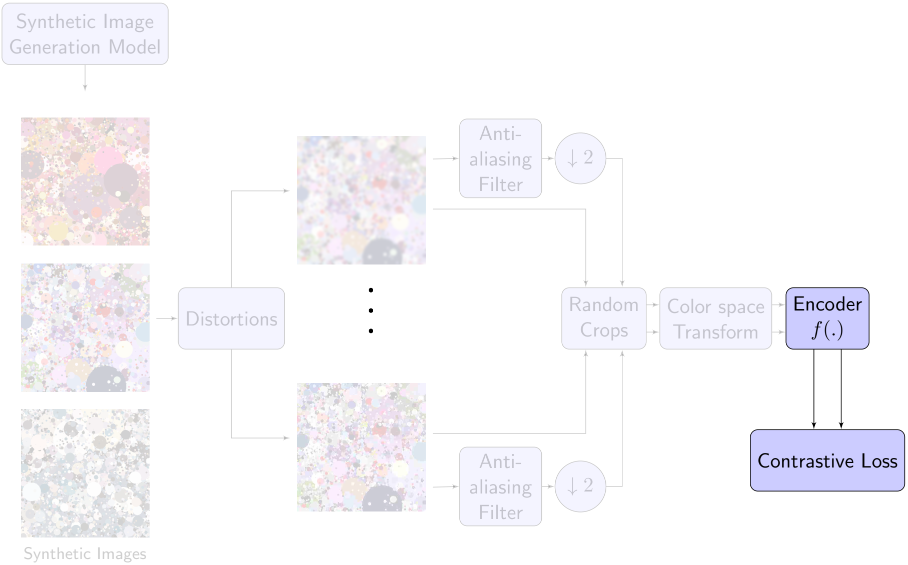
Data Augmentations/Transformations



Data Augmentations/Transformations

- Multiscale learning
  - Each image : two scales - full and half
  - Captures local and global image characteristics
- Colorspace transformations
  - Colorspaces - complementary artifacts information
  - RGB, LAB, HSV and grayscale
- Random cropping
  - Facilitates training - images of same size in a batch
  - Cropping - performed on both scales
  - Cropped versions - inherit distortion class labels of parent image

Contrastive Learning



Contrastive Loss

- Transformed and cropped images - encoder input
- Encoder : any CNN architecture - Resnet, VGG etc.
- Loss : normalized temperature-scaled cross entropy (NT-Xent)

$$\mathcal{L}_i = \frac{1}{|P(i)|} \sum_{j \in P(i)} -\log \frac{\exp(\phi(z_i, z_j)/\tau)}{\sum_{m=1}^{N} \mathbb{1}_{m \neq i} \exp(\phi(z_i, z_m)/\tau)}$$

  N - number of images present in the batch, $\mathbb{1}$ - indicator function, τ - temperature parameter, $P(i)$ - set containing image indices belonging to class of s_i, $|P(i)|$ - cardinality of $P(i)$

- Loss computation - pairwise between images in a batch

Evaluating Representations

- Evaluation - real images with distortions
- No additional fine-tuning on real data
- Correlation with human judgements - proxy for efficiency of features
- Mapping - regularized linear regression

$$y = Wk, \quad W^* = \underset{W}{\operatorname{argmin}} \sum_{i=1}^{N} (GT_i - y_i)^2 + \lambda \sum_{j=1}^{M} W_j^2,$$

  y - predicted scores, GT - ground-truth quality scores
  λ - regularization parameter, W - trainable weight vector
- Inference - no data transformations
  - Features at two scales - concatenation
- Evaluation metric - Spearman's rank order correlation (SROCC)

Outline of the Talk

Correlation with Human Judgments

Model	LIVE-IQA	CSIQ-IQA	TID	KADID
BRISQUE	0.939	0.746	0.604	0.528
CORNIA	0.947	0.678	0.678	0.516
HOSA	0.946	0.741	0.735	0.618
DB-CNN	**0.968**	**0.946**	0.816	0.851
Hyper-IQA	**0.962**	0.923	**0.840**	**0.852**
CONTRIQUE	0.960	**0.942**	**0.843**	**0.934**
Dead Leaves	0.940	0.852	0.703	0.776
Textured DL	0.950	0.920	0.751	0.820
Danbooru	0.960	**0.942**	0.790	0.910

- Training with textured DL - always improves performance
- Performance delta - reflect domain gap, relatively low on LIVE and CSIQ datasets

Correlation with Human Judgments

Model	LIVE-IQA	CSIQ-IQA	TID	KADID
BRISQUE	0.939	0.746	0.604	0.528
CORNIA	0.947	0.678	0.678	0.516
HOSA	0.946	0.741	0.735	0.618
DB-CNN	**0.968**	**0.946**	0.816	0.851
Hyper-IQA	**0.962**	0.923	**0.840**	**0.852**
CONTRIQUE	0.960	**0.942**	**0.843**	**0.934**
Dead Leaves	0.940	0.852	0.703	0.776
Textured DL	0.950	0.920	0.751	0.820
Danbooru	0.960	**0.942**	0.790	0.910

- Models trained on synthetic data - performance superior to traditional models
- Semantic information - significant, models trained on anime images perform better than DL

Shortcomings - Realistic Distortions

Method	KonIQ	CLIVE	FLIVE	SPAQ
BRISQUE	0.665	0.608	0.288	0.809
CORNIA	0.780	0.629	-	0.709
HOSA	0.805	0.640	-	0.846
DB-CNN	0.875	**0.851**	**0.554**	0.911
HyperIQA	**0.906**	**0.859**	0.535	**0.916**
CONTRIQUE	**0.894**	0.845	**0.580**	**0.914**
Dead Leaves	0.812	0.671	0.460	0.870
Textured Dead Leaves	0.820	0.677	0.485	0.872
Danbooru	0.841	0.715	0.520	0.886

- No authentic distortions in training data - poor performance on UGC
- Semantic information - greater significance in capturing realistic artifacts

Outline of the Talk

Conclusion and Future Work

- Investigation - synthetic data in unsupervised setting
- Analysis : effect of texture and semantic information
- Drawbacks : capturing realistic distortions
- Future Work
  - Training data - single image corrupted with multiple distortion types

Acknowledgements

- Collaborators

Dr. Neil (YouTube)

Dr. Yilin (YouTube)

Dr. Balu (YouTube)

Dr. Bovik (UT Austin)

- This work was supported by YouTube.

References

Pavan C Madhusudana, Neil Birkbeck, Yilin Wang, Balu Adsumilli, and Alan C Bovik. Image quality assessment using contrastive learning. arXiv preprint arXiv:2110.13266,2021.

Weixia Zhang, Kede Ma, Jia Yan, Dexiang Deng, and Zhou Wang. Blind image quality assessment using a deep bilinear convolutional neural network. IEEE Trans. Circuits Syst. Video Technol., 30(1):36–47, 2018.

Shaolin Su, Qingsen Yan, Yu Zhu, Cheng Zhang, Xin Ge,Jinqiu Sun, and Yanning Zhang. Blindly assess image qualityin the wild guided by a self-adaptive hyper network. In Proc. IEEE Conf. Comput. Vision Pattern Recognit., pages 3667–3676, 2020.

Thank You!