AEIMPS: Deep Autoencoder for Image Retargeting Quality Assessment

Levi C. Carvalho
PPGCC – IFCE
Federal Institute of Ceará (IFCE)
leviccxd@gmail.com

Saulo A. F. Oliveira
PPGCC – IFCE
Federal Institute of Ceará (IFCE)
saulo.oliveira@ifce.edu.br

***Abstract*—Evaluating retargeting image operators is a subjective task and, therefore, challenging to execute without human interference. Image Retargeting Quality Algorithms execute this task, giving some score to the retargeted image and, usually, trying to get a result similar to a human opinion since humans generally agree with each other on the quality of a resized image. Therefore, we propose an Autoencoder-based IRQA named AutoEncoder Information MaP Similarity (AEIMPS) to address this task using the NVAE architecture. In our experiments, besides the retargeting ratio, we use the latent space and the reconstructed image in the IRQA. AIEMPS achieved an average performance compared to other IRQAs in the literature.**

I. Introduction

Many multimedia applications use images, and retargeting images is necessary since the available space does not support the image dimension. In this context, content-aware image retargeting operators are employed [1], [2] because they can preserve the regions of interest and produce images with fewer distortions, see Fig. 1.

Although there exist a few algorithms [3]–[8], and due to the subjective nature of the problem, evaluating such algorithm results is not a trivial task. The main drawback is the different shapes between the original image and the retargeted one. Such a drawback cuts off many well-known image quality algorithms, such as Peak Signal-to-Noise-Ratio (PSNR), the Structural Similarity Index (SSIM) [9], and the Visual Information Fidelity Index (VIF) [10].

(a) Original. (b) Cropped. (c) Shift-map [5].

Fig. 1. Image retargeting process.

Many authors have proposed Image Retargeting Quality Algorithms (IRQAs) to overcome such a drawback [11]–[14]. Usually, such algorithms create a pixel matching mapping between the original image to the retargeting one, indicating a degree of content preservation [15]. Then, after applying some similarity criterion or measure of distance [16], one can recall the retargeting quality based on content matching and, perhaps, content relevance.

In this work, we propose an IRQA using the NVAE architecture [17], an Autoencoder (AE). Our proposal uses encoders and decoders, internal parts of NVAE, to predict quality in the sense of consumer perceptions. The motivation behind NVAE it is its ability to reconstruct the input (in our context, images) based on more concise representations. We named our proposal the **A**uto**E**ncoder **I**nformation **MaP** **S**imilarity (AEIMPS). From that, we can analyze the absence of visual distortion that may arise from resizing, which may compromise the quality of the final result. So far, to our knowledge, no studies have been produced that address the problem from that perspective.

This paper follows the subsequent organization. In Section II, we detail our proposal. Then, in Section III, we describe the experiments and results. In Section IV, we detail some possible shortcomings. Finally, in Section V, we discuss some final considerations.

II. Proposal

In short, our proposal combines both the latent space and the reconstructed image to compute the quality in a metric fusion fashion, see Fig. 2. We denote the latent space as $e(\cdot)$ and the reconstructed image as $d(e(\cdot))$. Additionally, we denote the retargeting factor as $\rho(\cdot,\cdot)$ – the aspect ratio difference between original $\mathbf{I}$ and retargeted $\mathbf{J}$ images.

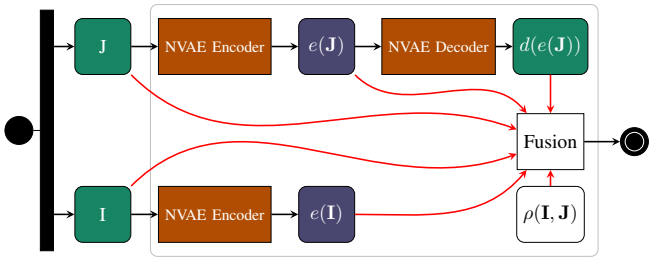


Fig. 2. AEIMPS flowchart.

We expect to observe two aspects by analyzing the latent space and the reconstructed image in AEIMPS. The first one is related to the latent space. We genuinely believe the code

(latent space vector) carries valuable features representing the image content. We support that a satisfactory retargeting image has a similar code to its original, while an unsatisfactory one has a code "missing" such valuable features. Thus, the difference between codes is how we account for such a quality. The second aspect AEIMPS aims for is the reconstruction effect of NVAE. It is expected that the reconstructed image drops some distortion, thus, being "restored", i.e., $d(e(\mathbf{J})) \approx \mathbf{I}$.

As for the fusion strategy, since more than one information is yielded $(e(\mathbf{J}), e(\mathbf{I}), d(e(\mathbf{J})), \rho(\mathbf{I}, \mathbf{J}))$, a successful strategy must combine such informations into a single score. This task has several fusion combination strategies such as simple averaging, product, linear addition, and non-linear combination. In Section III, we describe the basis scores and their combination.

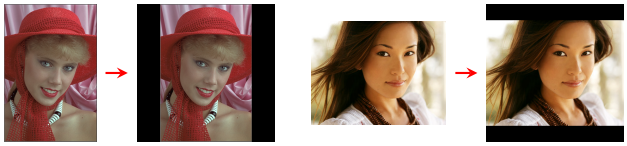
III. Experiments and Discussion

This section presents the experimental framework followed in this paper alongside the results and discussions on them. A GitHub repository with all code and experiments is available at https://github.com/LeviCC8/SIBGRAPI-WIP-AEIMPS.

A. Experiment Setup

1) Dataset: The experiments use the NTHU Retargeting Image Dataset (NRID). Such a dataset has 57 images with different shapes and contents in which each one is retargeted to 50% or 75% of the original shape size, using only three out of ten operators. Each pair of original and retargeted images is associated with a Mean Opinion Score (MOS), representing a subjective measure, ranging from 0 to 100, of how well the retargeting operator achieved a good result.

2) Pre-processing: We scaled each image pixel between 0 and 1 and resized them with padding to shape $3\times32\times32$ using the bilinear method[1], see Fig. 3. This resizing is necessary for the input size of the chosen AE architecture. We used the NVAE as AE, while its weights and configuration were taken from the official implementation at GitHub, choosing the one trained on ImageNet.



(a) Vertical padding resizing. (b) Horizontal padded resizing.

Fig. 3. Resizing for NVAE input layer.

3) Metric fusion: For the sake of simplicity, consider both $\mathbf{X}, \mathbf{Y}$ are tensors of rank-3 (rows, columns, and channels) such as $\mathbf{X} \in \mathbb{R}^{M\times N\times C}$ and $\mathbf{Y} \in \mathbb{R}^{M'\times N'\times C}$. Next, we describe the basis information and their combination.

$$\mathrm{cmae}(\mathbf{X}, \mathbf{Y}) = 1 - \frac{\| \bar{\mathbf{X}} - \bar{\mathbf{Y}} \|_1}{M \times N \times C}. \quad (1)$$

$$\mathrm{fnorm}(\mathbf{X}) = \| \mathbf{X} \|_{\mathcal{F}} . \quad (2)$$

$$\mathrm{cnorm}(\mathbf{X}) = \frac{1}{C}\sum_{c=1}^{C} \mathrm{fnorm}(\mathbf{X}_{*,*,c}). \quad (3)$$

$$\rho(\mathbf{X}, \mathbf{Y}) = \frac{\min(M, M')\min(N, N')}{\max(M, M')\max(N, N')}. \quad (4)$$

Also, the $\bar{\mathbf{X}}$ and $\bar{\mathbf{Y}}$ stand for their scaled version between 0 and 1, i.e., $0 \geq X_{*,*,*}, Y_{*,*,*} \geq 1$.

Regarding the derived fusion strategies, we detail them as follows:

- **FUSION 01**. The Compl. of Mean Absolute Error score between the image and its retargeted reconstruction from the AE, i.e., $\mathrm{cmae}(\mathbf{I}, d(e(\mathbf{J})))$;
- **FUSION 02**. Weighted Compl. of Mean Absolute Error by retargeting factor between the image and its retargeted reconstruction, i.e., $\rho(\mathbf{I}, \mathbf{J}) \cdot \mathrm{cmae}(\mathbf{I}, d(e(\mathbf{J})))$;
- **FUSION 03**. Weighted Compl. of Mean Absolute Error by the inverse of Retargeting factor between the image and its retargeted reconstruction, i.e. $\rho(\mathbf{I}, \mathbf{J})^{-1}\mathrm{cmae}(\mathbf{I}, d(e(\mathbf{J})))$;
- **FUSION 04**. The Compl. of Mean Absolute Error score between the image and its reconstruction from the AE to the power of the Retargeting factor, i.e., $\mathrm{cmae}(\mathbf{I}, d(e(\mathbf{J})))^{\rho(\mathbf{I},\mathbf{J})}$;
- **FUSION 05**. Retargeting factor to the power of the Compl. of Mean Absolute Error score between the image and its reconstruction from the AE, i.e., $\rho(\mathbf{I}, \mathbf{J})^{\mathrm{cmae}(\mathbf{I}, d(e(\mathbf{J})))}$;
- **FUSION 06 up to 09**. Linear combination (using α, β) between the Compl. of Mean Absolute Error score between the image and its reconstruction from the AE and Retargeting factor, i.e., $\alpha \cdot \mathrm{cmae}(\mathbf{I}, d(e(\mathbf{J}))) + \beta \cdot \rho(\mathbf{I}, \mathbf{J})$. Fusion 6 with $(0, 5, 0.6)$, Fusion 7 with $(0.8, 0.2)$, Fusion 8 with $(0.2, 0.8)$, and Fusion 9 with $(0.1, 0.9)$;
- **FUSION 10**. The Compl. of Mean Absolute Error score between the retargeted and its retargeted reconstruction from the AE, i.e., $\mathrm{cmae}(\mathbf{J}, d(e(\mathbf{J})))$.
- **FUSION 11**. Average norm by channel between the difference of original and retargeted image codes, i.e., $\mathrm{cnorm}(e(\mathbf{I}) - e(\mathbf{J}))$.
- **FUSION 12**. Weighted Average norm by channel between the difference of original and retargeted image codes by the Retargeting factor, i.e., $\rho(\mathbf{I}, \mathbf{J}) \cdot \mathrm{cnorm}(e(\mathbf{I}) - e(\mathbf{J}))$.
- **FUSION 13**. Norm of the retargeting code, i.e., $\mathrm{norm}(e(\mathbf{J}))$.

B. Results and Discussion

Since critical subjective components exist in the evaluation, one must employ a subjective test to assess how the IRQA output agrees with a typical human subject. To do so, we

[1]The NVAE is a AE based on PyTorch, i.e., it uses the channel-width-height notation. However, in this paper we use the width-height-channel notation.

employ four statistical measurements as suggested by the video quality experts group (VQEG) HDTV, a strategy also applied by [18]–[20].

The measurements are the linear correlation coefficient (LCC), the Spearman rank-order correlation coefficient (SRCC), the root mean square prediction error (RMSE), and the outlier ratio (OR). In all cases, higher LCC and SRCC coefficients represent higher agreement, while lower RMSE and OR values indicate smaller errors between the two scores; therefore, a better performance.

For a proper IRQA assessment, one must map each score according to the following five-parameter logistic function

$$f(x) = \beta_1 \left(\frac{1}{2} - \frac{1}{\exp\left(\beta_2 \left(x - \beta_3\right)\right)} \right) + \beta_4\, x + \beta_5, \qquad (5)$$

where $\{\beta_i\}_{i=1}^{5}$ are regression model parameters, determined by minimizing the sum of squared differences between IRQA scores and MOS.

In Table I, we show the results from the validation between the mapped scores and the MOS. The best score among the fusion is the linear combinations regarding the results. Such a finding can indicate some relevance in the $\rho(.)$ to compute the scores to measure the retargeting quality. Regarding the OR values in Table I, we observed some stability with 0. This finding reinforces that none of the combinations yielded a single atypical prediction. However, we noticed high RMSE between the MOS values and the scores.

In Table II, we present a comparison between our best result and other IRQAs in the literature. The other IRQAs scores were taken directly from [15]. Concerning LLC, SRCC, and RMSE, AEIMPS achieved an intermediate rank compared to the others. However, it achieved the best OR value which supports that our proposal is very promising.

TABLE I
RESULTS

IRQA	LLC	SRCC	RMSE	OR
Fusion 01	0.3404	0.3175	12.6944	0.0
Fusion 02	0.5288	0.4544	11.4587	0.0
Fusion 03	0.4313	0.3598	12.1809	0.0
Fusion 04	0.0436	0.0563	13.4880	0.0
Fusion 05	0.4963	0.3584	11.7205	0.0
Fusion 06	0.5268	0.4546	11.4756	0.0
Fusion 07	0.4476	0.4225	12.0730	0.0
Fusion 08	**0.5356**	**0.4557**	**11.4007**	**0.0**
Fusion 09	0.5333	0.4552	11.4211	0.0
Fusion 10	0.1390	0.1190	13.3699	0.0
Fusion 11	0.3314	0.3209	12.7382	0.0
Fusion 12	0.1250	0.1154	13.3972	0.0
Fusion 13	0.3774	0.4216	12.5023	0.0

We expect to observe two behaviors by analyzing the latent space and the reconstructed image. The first one, hopefully, the latent space, used to reconstruct the image in the AE output, carries valuable features to represent the image content. In the second one, we expect that the reconstructed retargeted image by the latent space drops some retargeting failures. Therefore, they could be used to measure the retargeting quality of operators computing the scores values.

(a) Face image for AEIMPS. (b) Kodim04 image for AEIMPS.

Fig. 4. AEIMPS in action.

TABLE II
PERFORMANCE COMPARISON BETWEEN IRQAS AND OURS (AEIMPS).

IRQA	LLC	SRCC	RMSE	OR
FMID (2017) [21]	0.7974	0.7984	8.3780	0.0643
ARS (2016) [14]	0.6835	0.6693	9.8550	0.0702
BIMS (SIM - NLIN) [15]	0.6503	0.6283	10.0484	0.0135
PGD (2014) [13]	0.5403	0.5409	11.3610	0.1520
GIST (2015) [22] [23]	0.5443	0.5114	11.3260	0.1579
NR (2016) [24]	0.5371	0.4926	-	0.1928
AEIMPS	**0.5356**	**0.4557**	**11.4007**	**0.0000**
GLS (2014) [12]	0.4622	0.4760	10.9320	0.1345
CSim (2011) [20]	0.4374	0.4662	12.1410	0.1520
EH (2001) [25]	0.3422	0.3288	12.6860	0.2047
EMD (2009) [26]	0.2760	0.2904	12.9770	0.1696
SIFT-Flow (2011) [11]	0.3141	0.2899	12.8170	0.1462
BDS (2008) [27]	0.2896	0.2887	12.9220	0.2164
PHOW (2007) [28]	0.3706	0.2308	12.5400	0.2222

IV. SHORTCOMINGS

In this section, we detail some shortcomings that may be harming the performance of AEIMPS. They are chosen based on some suspicions and are not proven. We describe the suspicious shortcomings:

- **Employed AE**. We use the NVAE architecture trained in the ImageNet dataset. Other architectures may have emphasis on the quality of the generated latent space such as InfoVAE [29]. Therefore, the latent space and the reconstructed image by NVAE may not be suitable enough to our problem;
- **Unsuitable fusion strategies**. In the fusion step of AEIMPS, we use the metrics based on the retargeting reduction factor, CMAE, and norm distance. Some of these may not be appropriate to measure the image retargeting quality;
- **Input shape of AE**. The input shape of the chosen architecture is $3 \times 32 \times 32$. Therefore, the images have to be resized to this shape, resulting in distortions and information loss. Thus, harming the final score.

V. CONCLUSION

In this work, we proposed an ANN-based IRQA using the NVAE architecture named **A**uto**E**ncoder **I**nformation **MaP** **S**imilarity (AEIMPS). AEIMPS employs scores based on the latent space and the reconstructed image information in several configurations, applying functions such as cmae and the norm of difference. The results were adjusted using a mapping function and compared with the MOS values from the dataset.

Statistical metrics (LLC, SRCC, RMSE, and OR) evaluated the results, indicating that a linear combination with the cmae score between the original image and its reconstruction from the NVAE and the retargeting factor achieved is our best configuration. Such a configuration ranked mid rank against other IRQA and avoided outliers during image retargeting quality prediction. As future work, we are investigating more basis metrics to explain how AEIMPS (framework) can perceive image retargeting quality alongside other AE architectures.

REFERENCES

[1] A. Shamir and O. Sorkine, "Visual media retargeting," in *ACM SIGGRAPH ASIA 2009 Courses*, ser. SIGGRAPH ASIA '09. New York, NY, USA: Association for Computing Machinery, 2009. [Online]. Available: https://doi.org/10.1145/1665817.1665828

[2] D. Vaquero, M. Turk, K. Pulli, M. Tico, and N. Gelfand, "A survey of image retargeting techniques," in *Applications of Digital Image Processing XXXIII*, vol. 7798. SPIE, 2010, pp. 328–342.

[3] M. Rubinstein, A. Shamir, and S. Avidan, "Improved seam carving for video retargeting," *ACM TRANSACTIONS ON GRAPHICS*, vol. 27, no. 3, AUG 2008, aCM SIGGRAPH Conference 2008, Singapore, SINGAPORE, AUG 11-15, 2008.

[4] Z. Karni, D. Freedman, and C. Gotsman, "Energy-based image deformation," *COMPUTER GRAPHICS FORUM*, vol. 28, no. 5, SI, pp. 1257–1268, JUL 2009, 7th Eurographics Symposium on Geometry Processing (SGP), Berlin, GERMANY, JUL 15-17, 2009.

[5] Y. Pritch, E. Kav-Venaki, and S. Peleg, "Shift-map image editing," in *2009 IEEE 12TH INTERNATIONAL CONFERENCE ON COMPUTER VISION (ICCV)*, ser. IEEE International Conference on Computer Vision. IEEE; IEEE Comp Soc, 2009, pp. 151–158, 12th IEEE International Conference on Computer Vision, Kyoto, JAPAN, SEP 29-OCT 02, 2009.

[6] M. Rubinstein, A. Shamir, and S. Avidan, "Multi-operator media retargeting," *ACM TRANSACTIONS ON GRAPHICS*, vol. 28, no. 3, AUG 2009, aCM SIGGRAPH Conference 2009, New Orleans, LA, 2009.

[7] L. Wolf, M. Guttmann, and D. Cohen-Or, "Non-homogeneous content-driven video-retargeting," in *2007 IEEE 11TH INTERNATIONAL CONFERENCE ON COMPUTER VISION, VOLS 1-6*, ser. IEEE International Conference on Computer Vision. IEEE, 2007, pp. 1418–1423, 11th IEEE International Conference on Computer Vision, Rio de Janeiro, BRAZIL, OCT 14-21, 2007.

[8] Y.-S. Wang, C.-L. Tai, O. Sorkine, and T.-Y. Lee, "Optimized scale-and-stretch for image resizing," *ACM TRANSACTIONS ON GRAPHICS*, vol. 27, no. 5, DEC 2008, aCM SIGGRAPH Conference 2008, Singapore, SINGAPORE, AUG 11-15, 2008.

[9] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE TRANSACTIONS ON IMAGE PROCESSING*, vol. 13, no. 4, pp. 600–612, APR 2004.

[10] H. Sheikh and A. Bovik, "Image information and visual quality," in *2004 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, VOL III, PROCEEDINGS: IMAGE AND MULTIDIMENSIONAL SIGNAL PROCESSING SPECIAL SESSIONS*. IEEE Signal Proc Soc; IEEE, 2004, pp. 709–712, iEEE International Conference on Acoustics, Speech, and Signal Processing, Montreal, CANADA, MAY 17-21, 2004.

[11] C. Liu, J. Yuen, and A. Torralba, "Sift flow: Dense correspondence across scenes and its applications," *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, vol. 33, no. 5, pp. 978–994, MAY 2011.

[12] J. Zhang and C. C. J. Kuo, "An objective quality of experience (qoe) assessment index for retargeted images," in *PROCEEDINGS OF THE 2014 ACM CONFERENCE ON MULTIMEDIA (MM'14)*. Assoc Comp Machinery; ACM SIGMM; FXPAL; Google; IBM; Microsoft Res; NASA Florida Space Grant Consortium; Yahoo Lab; Yandex, 2014, pp. 257–266, aCM Conference on Multimedia (MM), Univ Cent Florida, Orlando, FL, NOV 03-07, 2014.

[13] C.-C. Hsu, C.-W. Lin, Y. Fang, and W. Lin, "Objective quality assessment for image retargeting based on perceptual distortion and information loss," in *2013 IEEE INTERNATIONAL CONFERENCE ON VISUAL COMMUNICATIONS AND IMAGE PROCESSING (IEEE VCIP 2013)*. IEEE; Sarawak Convent Bur; Malaysia Convent & Exhibit Bur; IEEE Circuits & Syst Soc; Neuramatix, 2013, iEEE International Conference on Visual Communications and Image Processing (VCIP), Kuching, MALAYSIA, NOV 17-20, 2013.

[14] Y. Zhang, Y. Fang, W. Lin, X. Zhang, and L. Li, "Backward registration-based aspect ratio similarity for image retargeting quality assessment," *IEEE TRANSACTIONS ON IMAGE PROCESSING*, vol. 25, no. 9, pp. 4286–4297, SEP 2016.

[15] S. A. F. Oliveira, S. S. A. Alves, J. P. P. Gomes, and A. R. Rocha Neto, "A bi-directional evaluation-based approach for image retargeting quality assessment," *COMPUTER VISION AND IMAGE UNDERSTANDING*, vol. 168, no. SI, pp. 172–181, MAR 2018.

[16] A. Liu, W. Lin, H. Chen, and P. Zhang, "Image retargeting quality assessment based on support vector regression," *SIGNAL PROCESSING-IMAGE COMMUNICATION*, vol. 39, no. B, SI, pp. 444–456, NOV 2015.

[17] A. Vahdat and J. Kautz, "NVAE: A deep hierarchical variational autoencoder," in *Neural Information Processing Systems (NeurIPS)*, 2020.

[18] L. Ma, W. Lin, C. Deng, and K. N. Ngan, "Image retargeting quality assessment: A study of subjective scores and objective metrics," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 6, pp. 626–639, 2012.

[19] Y. Zhang, Y. Fang, W. Lin, X. Zhang, and L. Li, "Backward registration-based aspect ratio similarity for image retargeting quality assessment," *IEEE Transactions on Image Processing*, vol. 25, no. 9, pp. 4286–4297, 2016.

[20] Y.-J. Liu, X. Luo, Y.-M. Xuan, W.-F. Chen, and X.-L. Fu, "Image retargeting quality assessment," *COMPUTER GRAPHICS FORUM*, vol. 30, no. 2, pp. 583–592, 2011.

[21] Y. Zhang, K. N. Ngan, L. Ma, and H. Li, "Objective quality assessment of image retargeting by incorporating fidelity measures and inconsistency detection," *IEEE TRANSACTIONS ON IMAGE PROCESSING*, vol. 26, no. 12, pp. 5980–5993, DEC 2017.

[22] L. Ma, C. Deng, W. Lin, K. N. Ngan, and L. Xu, *Retargeted Image Quality Assessment: Current Progresses and Future Trends*. Cham: Springer International Publishing, 2015, pp. 213–242. [Online]. Available: https://doi.org/10.1007/978-3-319-10368-6_8

[23] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *INTERNATIONAL JOURNAL OF COMPUTER VISION*, vol. 42, no. 3, pp. 145–175, 2001.

[24] L. Ma, L. Xu, Y. Zhang, Y. Yan, and K. N. Ngan, "No-reference retargeted image quality assessment based on pairwise rank learning," *IEEE TRANSACTIONS ON MULTIMEDIA*, vol. 18, no. 11, pp. 2228–2237, NOV 2016.

[25] D. Messing, P. van Beek, and J. Errico, "The mpeg-7 colour structure descriptor: Image description using colour and local spatial information," in *2001 INTERNATIONAL CONFERENCE ON IMAGE PROCESSING, VOL I, PROCEEDINGS*, ser. IEEE International Conference on Image Processing ICIP. IEEE Signal Processing Soc; IEEE, 2001, pp. 670–673, international Conference on Image Processing (ICIP 2001), THESSALONIKI, GREECE, OCT 07-10, 2001.

[26] O. Pele and M. Werman, "Fast and robust earth mover's distances," in *2009 IEEE 12TH INTERNATIONAL CONFERENCE ON COMPUTER VISION (ICCV)*, ser. IEEE International Conference on Computer Vision. IEEE; IEEE Comp Soc, 2009, pp. 460–467, 12th IEEE International Conference on Computer Vision, Kyoto, JAPAN, SEP 29-OCT 02, 2009.

[27] D. Simakov, Y. Caspi, E. Shechtman, and M. Irani, "Summarizing visual data using bidirectional similarity," in *2008 IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, VOLS 1-12*, ser. IEEE Conference on Computer Vision and Pattern Recognition. IEEE Comp Soc, 2008, pp. 3887+, iEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, JUN 23-28, 2008.

[28] A. Bosch, A. Zisserman, and X. Munoz, "Image classification using random forests and ferns," in *2007 IEEE 11TH INTERNATIONAL CONFERENCE ON COMPUTER VISION, VOLS 1-6*, ser. IEEE International Conference on Computer Vision. IEEE, 2007, pp. 1863–1870, 11th IEEE International Conference on Computer Vision, Rio de Janeiro, BRAZIL, OCT 14-21, 2007.

[29] S. Zhao, J. Song, and S. Ermon, "Infovae: Balancing learning and inference in variational autoencoders," in *Proceedings of the aaai conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 5885–5892.