Multi-OctConv: Reducing Memory Requirements in Image Generative Adversarial Networks

Francisco Tobar Méndez

Dr. Claudio Torres

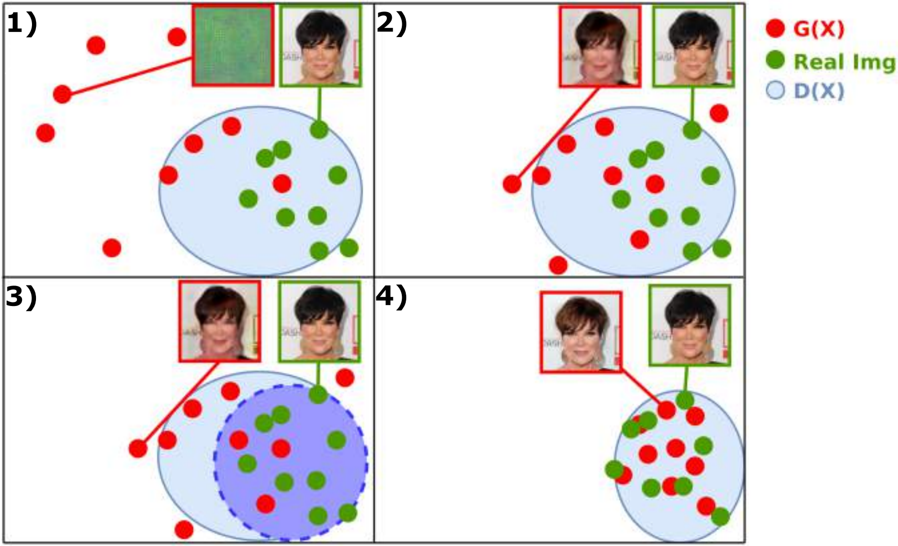
XXXIX International Conference of the Chilean Computer Science Society

Universidad Técnica Federico Santa María

Departamento de Informática

Content Table

Generative Adversarial Network



GANs for Image Generation

Model	Year	Features Control Methods
Cycle-GAN [1]	2017	Model-Feature
StarGAN [2]	2018	Parameter
StarGAN v2 [3]	2019	Parameter/Secondary Image
StyleGAN v2 [4]	2019	Parameter/Secondary Image
GANSpace v2 [5]	2020	Parameter/Secondary Image
StyleGANv2+ Distillation [6]	2020	Parameter/Secondary Image

- **Features Control Methods**: Methods to control outputs features.
- All these GAN use a *Image-to-Image* approach to generate the outputs images.
- All these GAN use convolutional encoder/decoder structures.

Samples of GANs Genrated Images

a) Cycle-GAN [1] b) StarGAN [2] c) StarGAN v2 [3]

d) StyleGAN v2 [4] e) GANSpace v2 [5] f) StyleGAN v2+Dist [6]

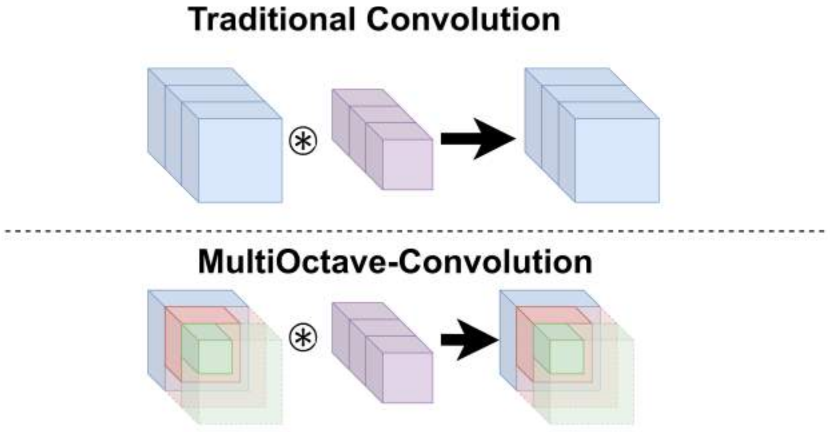
Memory Usage

- High amount of memory is required in the training phase due to convolutional encoder/decoder structures in GANs.
- High performance hardware are required to train this networks.

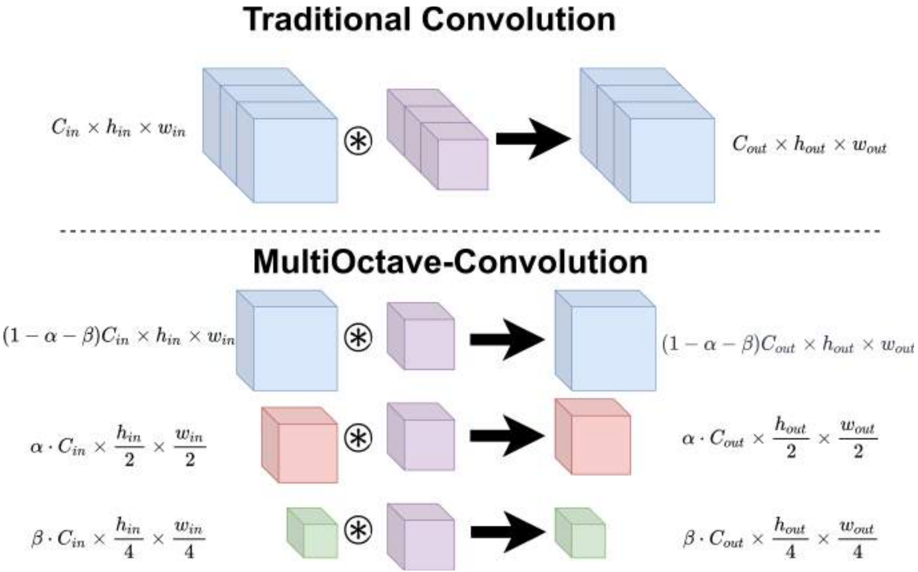
Convolution Memory Reduction Methods

- Memory reduction by removing elements of the model.
    - Pruning techniques [7, 8, 9].
    - Sparse convolution operators [10, 11, 12].
- Methods that reduce memory through the modifications of the filters of convolutional layers.
    - HetConv [13].
    - LeanConv [14].
- Techniques that reduce the volume of data processed by each layer.
    - Octave Convolution [15].

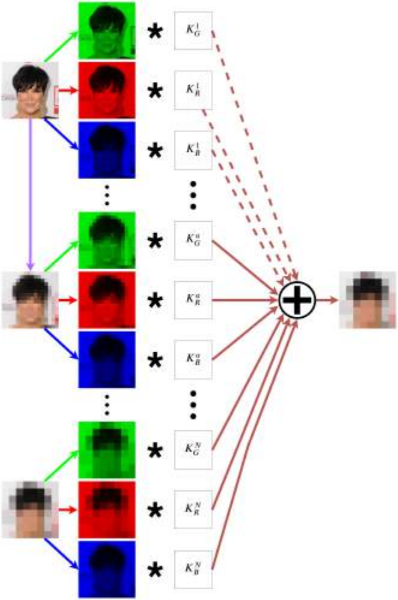
Multi-Octave Convolution (Multi-OctConv)



Multi-Octave Convolution (Multi-OctConv)

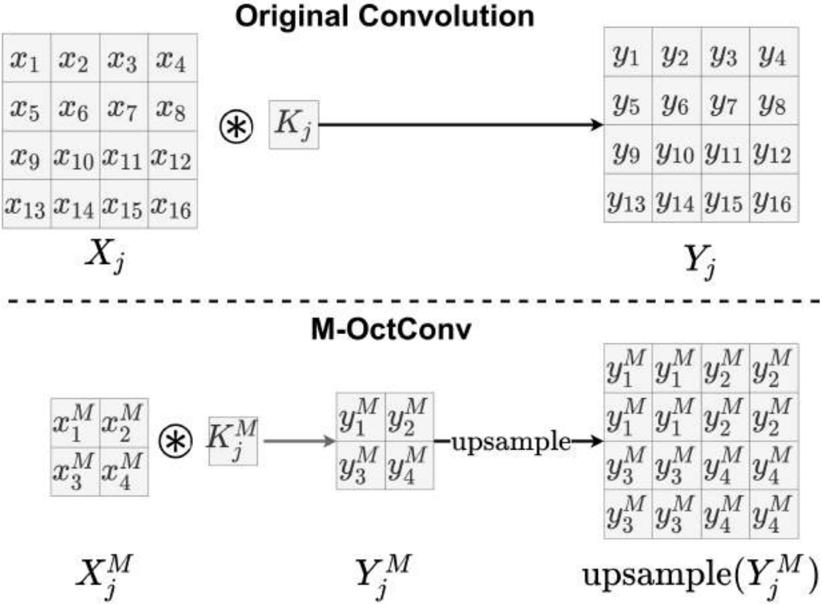


Multi-Octave Convolution (Multi-OctConv)

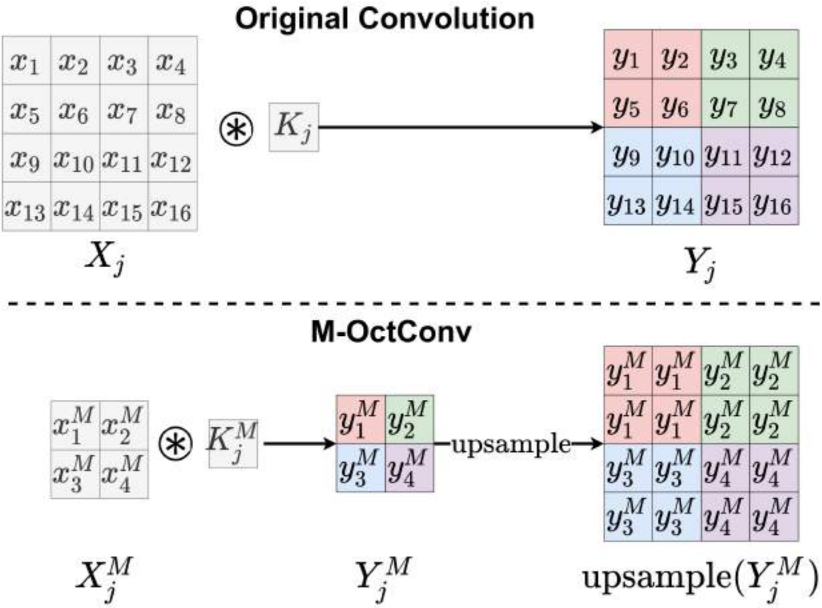


- Variation of OctConv with an extra frequency level
- Better memory reduction than octConv
- Skip some frequency communication.
- Reduce the number of parameters traditional convolutional layers

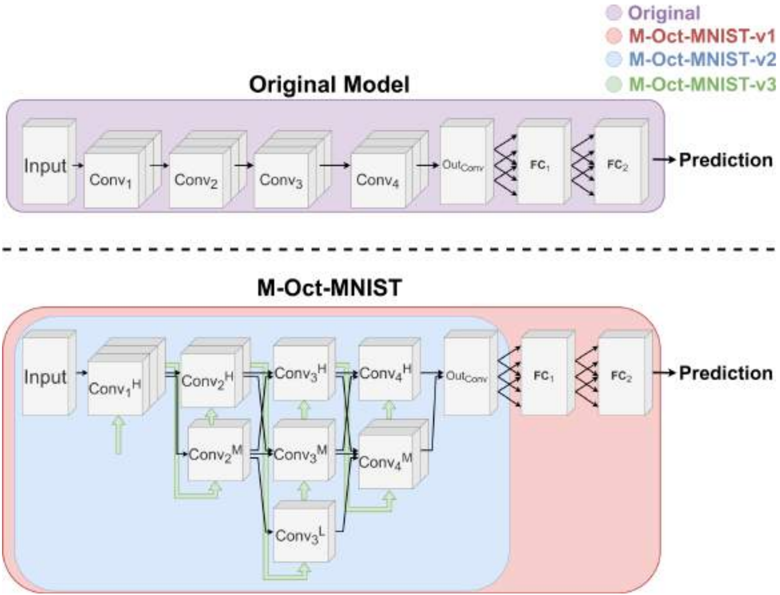
Induced Error



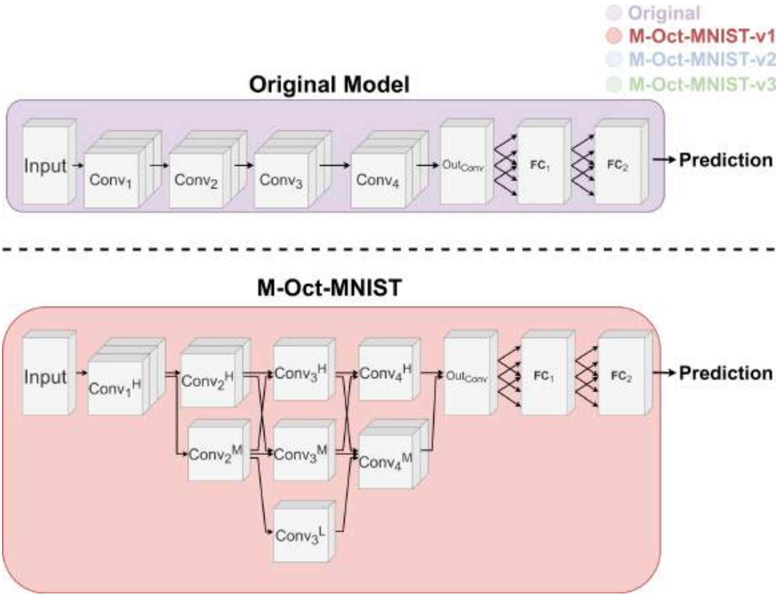
Induced Error



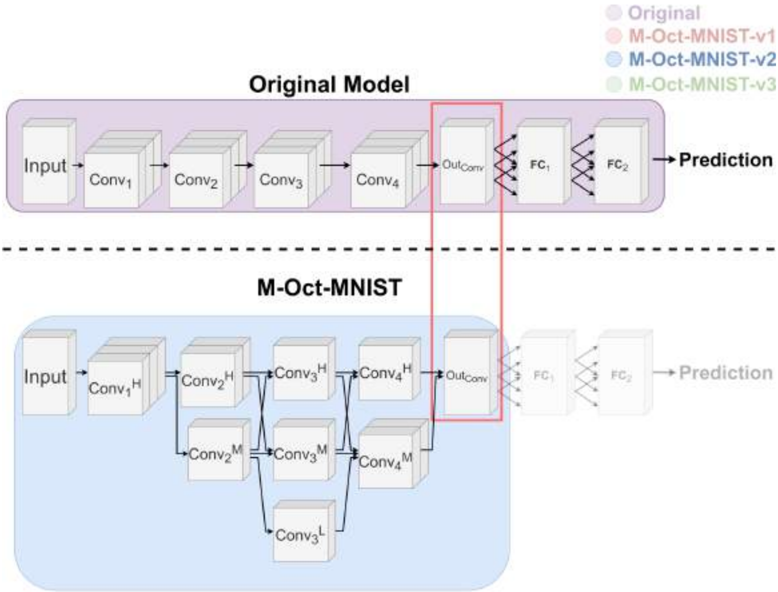
M-OctConv Error Analysis



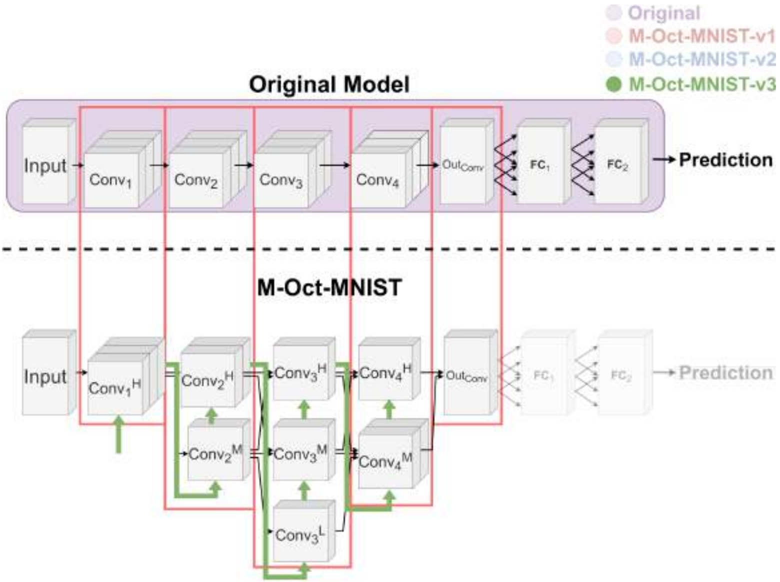
M-OctConv Error Analysis



M-OctConv Error Analysis



M-OctConv Error Analysis



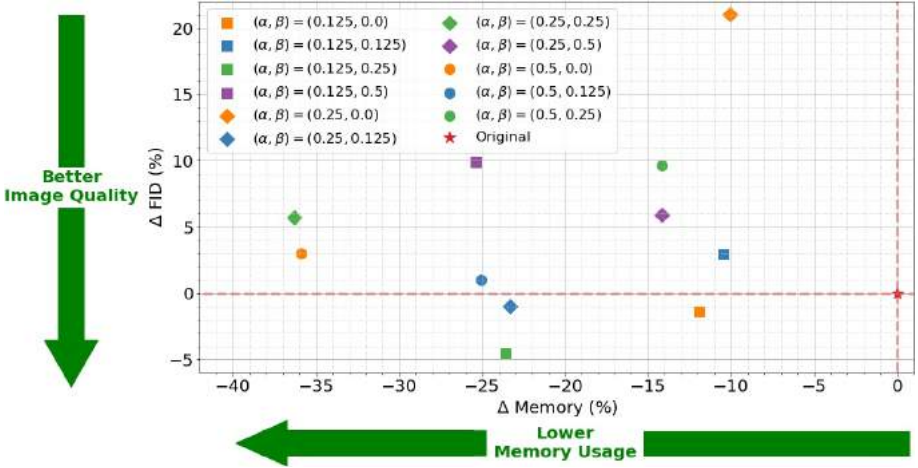
M-OctConv Error Analysis results

Model	Classification Accuracy	Output Reconstruction Mean Relative Error
Original	96.75 %	-
M-OctMNIST-v1	96.70 %	-
M-OctMNIST-v2	96.44 %	1.49e-08
M-OctMNIST-v3	94.19 %	1.37e-07

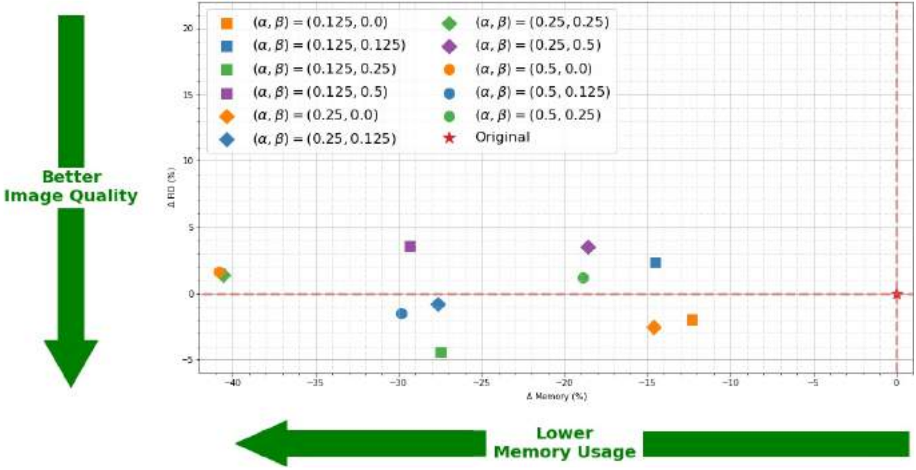
M-OctConv Evaluation Over StarGAN

- Training time in less than a day.
- Produce realistic images.
- Works over two Datasets:
    - CelebFaces Attributes (CelebA) dataset [16] (200.000 images)
    - Radboud Faces Database (RaFD) [17] (8.000 images)

M-OctConv Evaluation Over StarGAN CelebA



M-OctConv Evaluation Over StarGAN RaFD



M-OctConv Evaluation Over StarGAN

α	β	#Params		% Original	
		RafD	CelebA	RafD	CelebA
0.125	0.0	53.226.560	53.192.576	100 %	100 %
	0.125	50.383.936	50.349.952	94.66 %	94.66 %
	0.25	48.479.296	48.445.312	91.08 %	91.08 %
	0.5	47.483.968	47.449.984	89.21 %	89.20 %

M-OctConv Evaluation Over StarGAN Examples

a) StarGAN CelebaA b) Oct-StarGAN CelebaA c) M-Oct-StarGAN CelebaA

d) StarGAN RafD e) Oct-StarGAN RafD f) M-Oct-StarGAN RafD

Conclusions

With M-OctConv, we are able to:

- Reduce the memory required in the training process.
- Modify only the convolution layers of the models without changing the architecture.
- Keep a similar quality of the images generated by the model with M-OctConv and even slightly improving it in some cases.
- Use other methods for reduction of memory usage in convolutional layers alongside M-OctConv

Future Work

- Determine how many extra low levels can be added before damaging the quality of the convolution.
- Evaluate M-OctConv on other GAN models and other types of models

Acknowledgment

- *Dirección de Postgrado y Programas*, UTFSM, Chile.
- CCTVal CONICYT PIA/Basal FB0821,
- This research/thesis was partially supported by the supercomputing infrastructure of the NLHPC (ECM-02).

Multi-OctConv: Reducing Memory Requirements in Image Generative Adversarial Networks

Francisco Tobar Méndez

Dr. Claudio Torres

XXXIX International Conference of the Chilean Computer Science Society

Universidad Técnica Federico Santa María

Departamento de Informática

Referencias I

J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8789–8797.

Y. Choi, Y. Uh, J. Yoo, and J.-W. Ha, "Stargan v2: Diverse image synthesis for multiple domains," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 8188–8197.

T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of stylegan," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 8110–8119.

E. Härkönen, A. Hertzmann, J. Lehtinen, and S. Paris, "Ganspace: Discovering interpretable gan controls," *arXiv preprint arXiv:2004.02546*, 2020.

Y. Viazovetskyi, V. Ivashkin, and E. Kashin, "Stylegan2 distillation for feed-forward image manipulation," *arXiv preprint arXiv:2003.03581*, 2020.

Referencias II

P. Singh, V. S. R. Kadi, N. Verma, and V. P. Namboodiri, "Stability based filter pruning for accelerating deep cnns," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV).* IEEE, 2019, pp. 1166–1174.

P. Molchanov, S. Tyree, T. Karras, T. Aila, and J. Kautz, "Pruning convolutional neural networks for resource efficient inference," *arXiv preprint arXiv:1611.06440*, 2016.

F. Tung and G. Mori, "Clip-q: Deep network compression learning by in-parallel pruning-quantization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7873–7882.

W. Wen, C. Wu, Y. Wang, Y. Chen, and H. Li, "Learning structured sparsity in deep neural networks," in *Advances in neural information processing systems*, 2016, pp. 2074–2082.

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

S. Han, J. Pool, S. Narang, H. Mao, E. Gong, S. Tang, E. Elsen, P. Vajda, M. Paluri, J. Tran *et al.*, "Dsd: Dense-sparse-dense training for deep neural networks," *arXiv preprint arXiv:1607.04381*, 2016.

P. Singh, V. K. Verma, P. Rai, and V. P. Namboodiri, "Hetconv: Heterogeneous kernel-based convolutions for deep cnns," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4835–4844.

Referencias III

J. Ephrath, M. Eliasof, L. Ruthotto, E. Haber, and E. Treister, "Leanconvnets: Low-cost yet effective convolutional neural networks," *IEEE Journal of Selected Topics in Signal Processing*, 2020.

Y. Chen, H. Fang, B. Xu, Z. Yan, Y. Kalantidis, M. Rohrbach, S. Yan, and J. Feng, "Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution," *arXiv preprint arXiv:1904.05049*, 2019.

Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proceedings of International Conference on Computer Vision (ICCV)*, December 2015.

O. Langner, R. Dotsch, G. Bijlstra, D. H. Wigboldus, S. T. Hawk, and A. Van Knippenberg, "Presentation and validation of the radboud faces database," *Cognition and emotion*, vol. 24, no. 8, pp. 1377–1388, 2010.

I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein gan," *arXiv preprint arXiv:1701.07875*, 2017.

I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," in *Advances in neural information processing systems*, 2017, pp. 5767–5777.

X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. Paul Smolley, "Least squares generative adversarial networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2794–2802.

Referencias IV

X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, “On the effectiveness of least squares generative adversarial networks,” *IEEE transactions on pattern analysis and machine intelligence*, 2018.

M. Mirza and S. Osindero, “Conditional generative adversarial nets,” *arXiv preprint arXiv:1411.1784*, 2014.

A. Anoosheh, E. Agustsson, R. Timofte, and L. Van Gool, “Combogan: Unrestrained scalability for image domain translation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 783–790.

T. Xiao, J. Hong, and J. Ma, “Elegant: Exchanging latent encodings with gan for transferring multiple face attributes,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 168–184.

X. Huang, M.-Y. Liu, S. Belongie, and J. Kautz, “Multimodal unsupervised image-to-image translation,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 172–189.

Z. He, W. Zuo, M. Kan, S. Shan, and X. Chen, “Attgan: Facial attribute editing by only changing what you want,” *IEEE Transactions on Image Processing*, 2019.

H.-Y. Lee, H.-Y. Tseng, J.-B. Huang, M. Singh, and M.-H. Yang, “Diverse image-to-image translation via disentangled representations,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 35–51.

Referencias V

X. Yu, X. Cai, Z. Ying, T. Li, and G. Li, "Singlegan: Image-to-image translation by a single-generator network using multiple generative adversarial learning," in *Asian Conference on Computer Vision*. Springer, 2018, pp. 341–356.

D. Berthelot, T. Schumm, and L. Metz, "Began: Boundary equilibrium generative adversarial networks," *arXiv preprint arXiv:1703.10717*, 2017.

Y. Lu, Y.-W. Tai, and C.-K. Tang, "Attribute-guided face generation using conditional cyclegan," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 282–297.

R. Sun, C. Huang, J. Shi, and L. Ma, "Mask-aware photorealistic face attribute manipulation," *arXiv preprint arXiv:1804.08882*, 2018.

T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," in *Advances in neural information processing systems*, 2016, pp. 2234–2242.

M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," in *Advances in Neural Information Processing Systems*, 2017, pp. 6626–6637.

A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," 2017.

Referencias VI

A. Jolicoeur-Martineau, “The relativistic discriminator: a key element missing from standard gan,” *arXiv preprint arXiv:1807.00734*, 2018.

L.-C. Yang, S.-Y. Chou, and Y.-H. Yang, “Midinet: A convolutional generative adversarial network for symbolic-domain music generation,” *arXiv preprint arXiv:1703.10847*, 2017.

J. Wu, C. Zhang, T. Xue, B. Freeman, and J. Tenenbaum, “Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling,” in *Advances in neural information processing systems*, 2016, pp. 82–90.

H. Kameoka, T. Kaneko, K. Tanaka, and N. Hojo, “Stargan-vc: Non-parallel many-to-many voice conversion using star generative adversarial networks,” in *2018 IEEE Spoken Language Technology Workshop (SLT)*. IEEE, 2018, pp. 266–273.

L. de Oliveira, M. Paganini, and B. Nachman, “Learning particle physics by example: location-aware generative adversarial networks for physics synthesis,” *Computing and Software for Big Science*, vol. 1, no. 1, p. 4, 2017.

Y. Tu, Y. Lin, J. Wang, and J.-U. Kim, “Semi-supervised learning with generative adversarial networks on digital signal modulation classification,” *Comput. Mater. Continua*, vol. 55, no. 2, pp. 243–254, 2018.

Y. Chen, Y. Wang, D. Kirschen, and B. Zhang, “Model-free renewable scenario generation using generative adversarial networks,” *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 3265–3275, 2018.

Referencias VII

I. Sutskever, R. Jozefowicz, K. Gregor, D. Rezende, T. Lillicrap, and O. Vinyals, “Towards principled unsupervised learning,” *arXiv preprint arXiv:1511.06440*, 2015.

A. Pumarola, A. Agudo, A. M. Martinez, A. Sanfeliu, and F. Moreno-Noguer, “Ganimation: Anatomically-aware facial animation from a single image,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 818–833.

M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, “TensorFlow: Large-scale machine learning on heterogeneous systems,” 2015, software available from tensorflow.org. [Online]. Available: http://tensorflow.org/

F. Tobar and C. Torres. (2020) M-octconv repository. Accessed October 22, 2020. [Online]. Available: https://github.com/ftobarm/MultiOctConv

G.-J. Qi, “Loss-sensitive generative adversarial networks on lipschitz densities,” *International Journal of Computer Vision*, pp. 1–23, 2019.

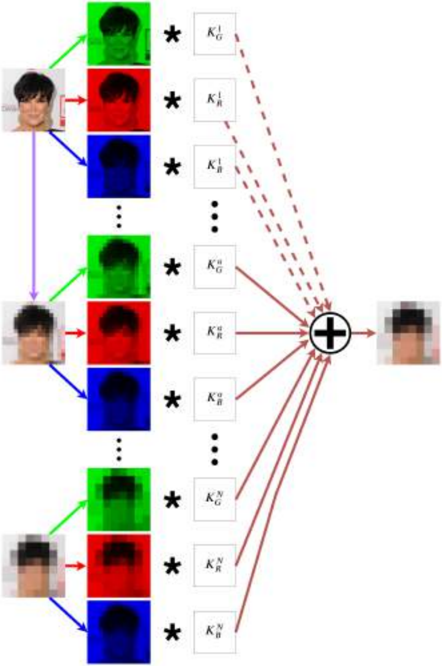
J.-H. Luo, H. Zhang, H.-Y. Zhou, C.-W. Xie, J. Wu, and W. Lin, “Thinet: pruning cnn filters for a thinner net,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 10, pp. 2525–2538, 2018.

Referencias VIII

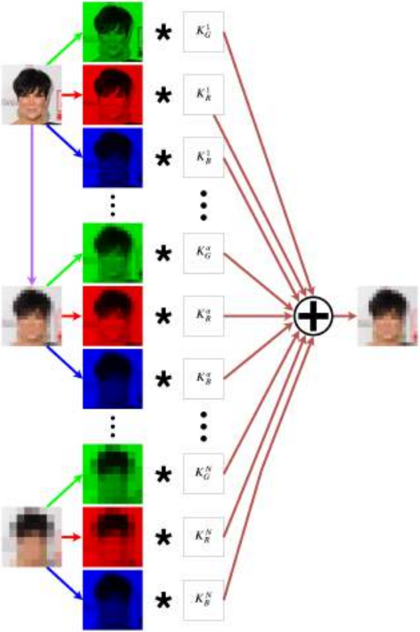
M. Arjovsky and L. Bottou, "Towards principled methods for training generative adversarial networks," *arXiv preprint arXiv:1701.04862*, 2017.

Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

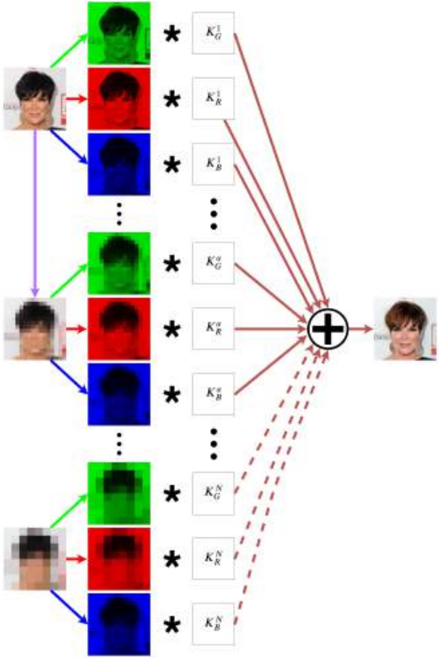
Multi-OctConv



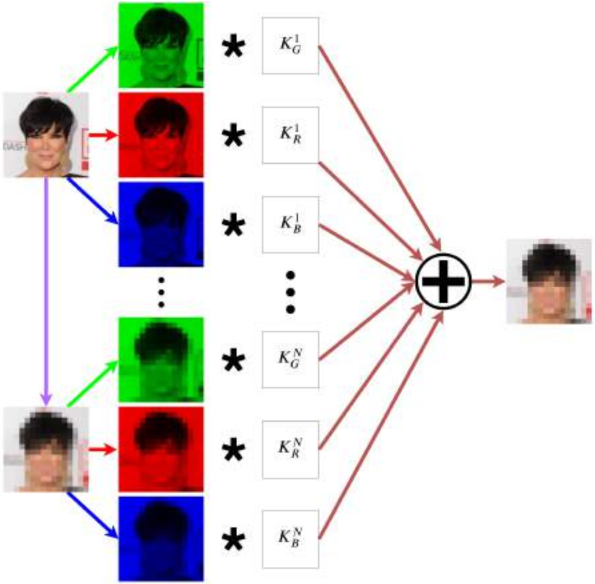
Multi-OctConv



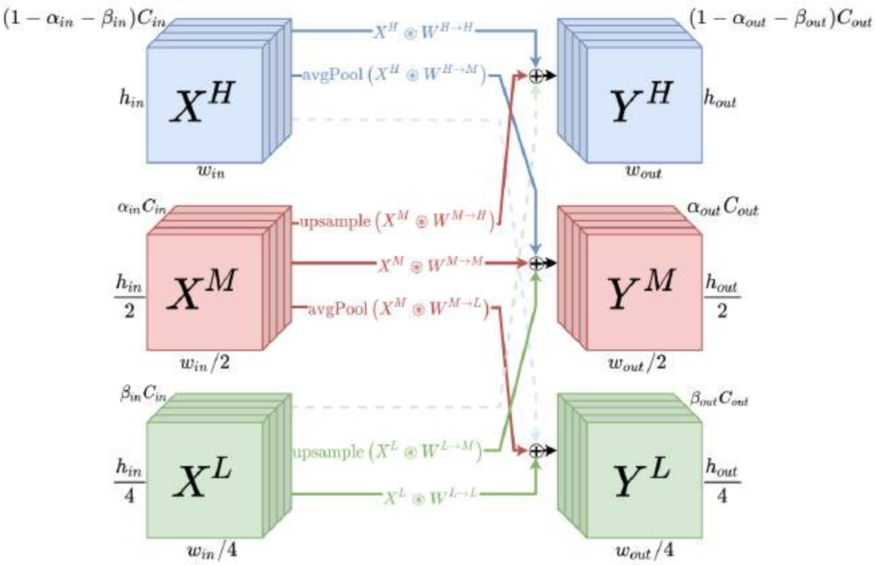
Multi-OctConv



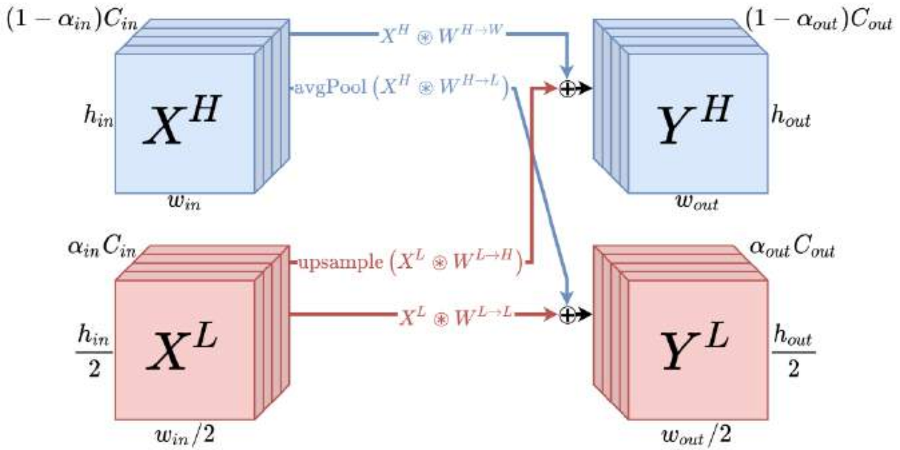
Multi-OctConv



Multi-OctConv



Multi-OctConv



Frequency level	Memory Required
High Frequency	$w_{in} \cdot h_{in} \cdot C_{in} \cdot (1 - \alpha_{in} - \beta_{in})$
Mid Frequency	$\frac{w_{in}}{2} \cdot \frac{h_{in}}{2} \cdot C_{in} \cdot \alpha_{in}$
Low Frequency	$\frac{w_{in}}{4} \cdot \frac{h_{in}}{4} \cdot C_{in} \cdot \beta_{in}$
Total	$w_{in} \cdot h_{in} \cdot C_{in} \cdot (1 - \frac{3}{4}\alpha_{in} - \frac{15}{16}\beta_{in})$

Frequency level	Convolution	# Parameters
High Frequency	$\text{Conv}_{H\rightarrow H}$	$C_{in} \cdot C_{out} \cdot k^2 \cdot (1 - \alpha_{in} - \beta_{in})^2$
	$\text{Conv}_{H\rightarrow M}$	$C_{in} \cdot C_{out} \cdot k^2 \cdot (1 - \alpha_{in} - \beta_{in}) \cdot \alpha_{in}$
Mid Frequency	$\text{Conv}_{M\rightarrow H}$	$C_{in} \cdot C_{out} \cdot k^2 \cdot \alpha_{in} \cdot (1 - \alpha_{in} - \beta_{in})$
	$\text{Conv}_{M\rightarrow M}$	$C_{in} \cdot C_{out} \cdot k^2 \cdot \alpha_{in}^2$
	$\text{Conv}_{M\rightarrow L}$	$C_{in} \cdot C_{out} \cdot k^2 \cdot \alpha_{in} \cdot \beta_{in}$
Low Frequency	$\text{Conv}_{L\rightarrow M}$	$C_{in} \cdot C_{out} \cdot k^2 \cdot \beta_{in} \cdot \alpha_{in}$
	$\text{Conv}_{L\rightarrow L}$	$C_{in} \cdot C_{out} \cdot k^2 \cdot \beta_{in}^2$
Total	-	$C_{in} \cdot C_{out} \cdot k^2 \cdot (1 + 2\beta_{in}(\alpha_{in} + \beta_{in} - 1))$