Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Network

Unsupervised Representation Learning

Learning reusable feature representations from large unlabeled datasets

Challenge

GAN is unstable to train

Contribution

Propose and evaluate a set of constraints on the architectural topology of Convolutional GANs that make them stable to train in most setting.

Architecture guidelines for stable Deep Convolutional GANs

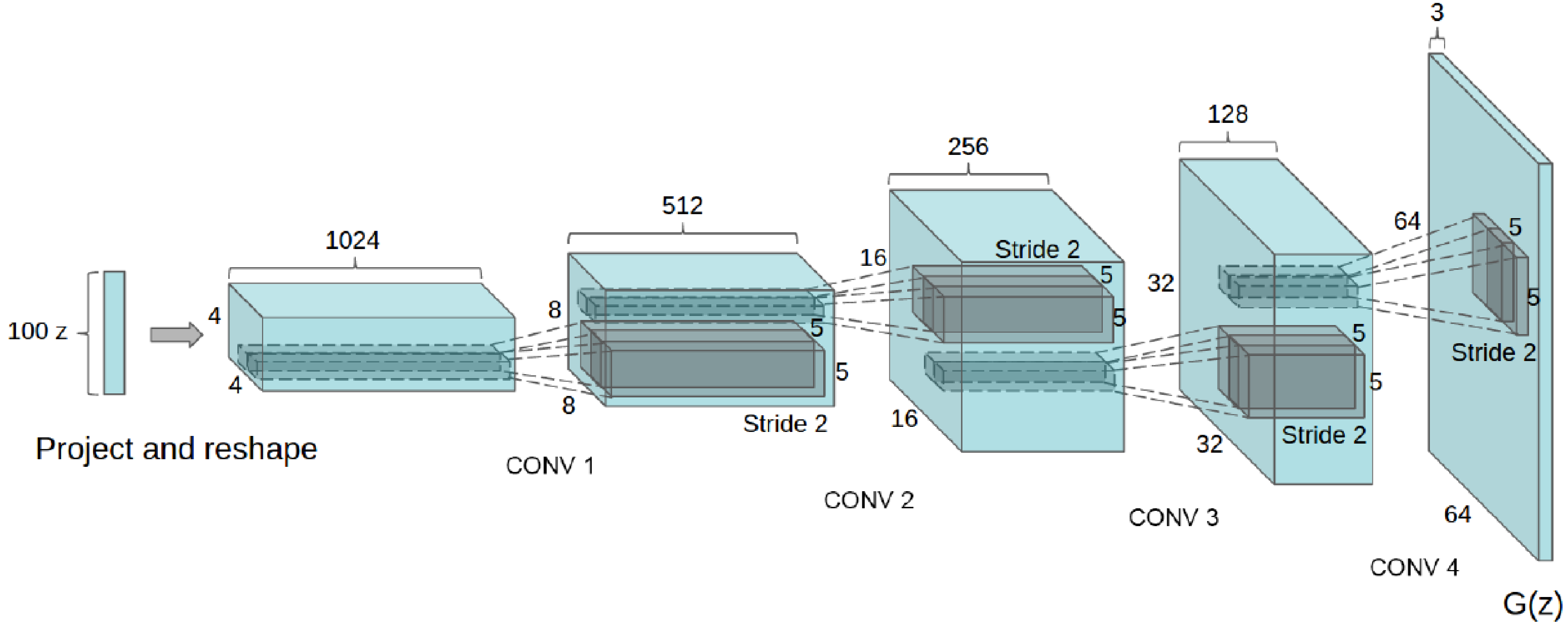
- Replace any pooling layers with strided convolutions (discriminator) and fractional-strided convolutions (generator).
- Use batchnorm in both the generator and the discriminator.
- Remove fully connected hidden layers for deeper architectures.
- Use ReLU activation in generator for all layers except for the output, which uses Tanh.
- Use LeakyReLU activation in the discriminator for all layers.

LeakyReLU

$$y_i = \begin{cases} x_i & \text{if } x_i \geq 0 \\ \frac{x_i}{a_i} & \text{if } x_i < 0, \end{cases}$$

where α is a fixed parameter in range $(1, +\infty)$

DCGAN generator used for LSUN scene modeling



Experiment

As a feature extractor

CIFAR-10 dataset

- Trained on Imagenet-1k
- Use the discriminator’s convolutional features from all layer with 4x4 max-pooling
- Train a regularized linear L2-SVM classifier

Model	Accuracy	Accuracy (400 per class)	max # of features units
1 Layer K-means	80.6%	63.7% (±0.7%)	4800
3 Layer K-means Learned RF	82.0%	70.7% (±0.7%)	3200
View Invariant K-means	81.9%	72.6% (±0.7%)	6400
Exemplar CNN	84.3%	77.4% (±0.2%)	1024
DCGAN (ours) + L2-SVM	82.8%	73.8% (±0.4%)	512

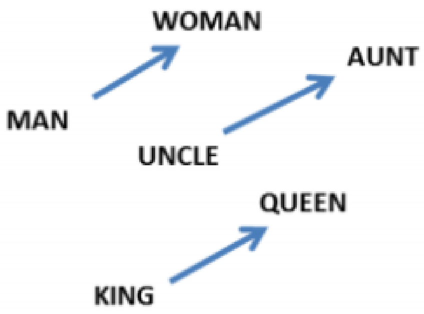
SVHN (StreetView House Numbers dataset)

Model	error rate
KNN	77.93%
TSVM	66.55%
M1+KNN	65.63%
M1+TSVM	54.33%
M1+M2	36.02%
SWWAE without dropout	27.83%
SWWAE with dropout	23.56%
DCGAN (ours) + L2-SVM	22.48%
Supervised CNN with the same architecture	28.87% (validation)

Random filters **Trained filters**

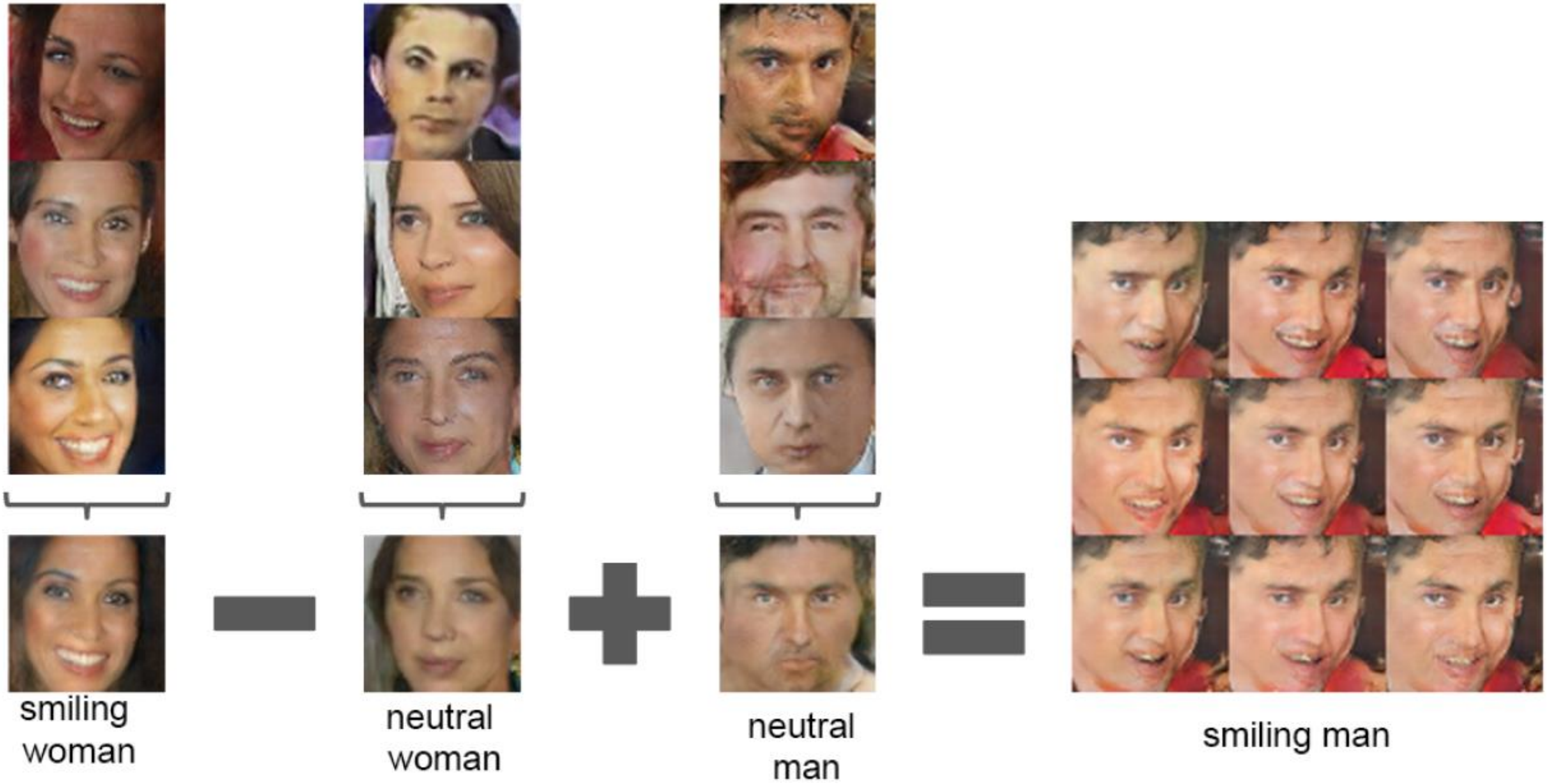
Vector arithmetic on face samples

Word Embedding



$$W(\text{“woman”}) - W(\text{“man”}) \simeq W(\text{“aunt”}) - W(\text{“uncle”})$$

$$W(\text{“woman”}) - W(\text{“man”}) \simeq W(\text{“queen”}) - W(\text{“king”})$$


smiling woman
neutral woman
neutral man
smiling man

References

Radford, Alec, L. Metz, and S. Chintala. "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks." *Computer Science* (2015).

Goodfellow, Ian J., et al. "Generative adversarial nets." *International Conference on Neural Information Processing Systems* MIT Press, 2014:2672-2680.

Springenberg, Jost Tobias, et al. "Striving for Simplicity: The All Convolutional Net." *Eprint Arxiv* (2014).

Zeiler, Matthew D., and R. Fergus. "Visualizing and Understanding Convolutional Networks." 8689(2014):818-833.

Optimization problem

$$\min_{W,\Sigma,P} \sum_{k=1}^{q}\|w_k\|^2 + C_1\sum_{i=1}^{m}\frac{1}{|\hat{Y}||\acute{Y}|}\sum_{(y_k,y_l)\in\hat{Y}\times\acute{Y}}\xi_{ikl} + C_2\sum_{i=1}^{m}\sum_{k\in\hat{Y}}P_{ik}\cdot l(f_k(x_i),-1)$$
$$-C_3\sum_{i=1}^{m}\sum_{k\in\hat{Y}_i}S_{k\cdot}(P_{ik}\cdot P_{i\cdot})^T - C_4\sum_{i=1}^{m}\sum_{k\in\hat{Y}_i}P_{ik}\cdot\|x_i - Q_k\|$$

Subject: $\langle w_k - w_l, x_i\rangle + b_k - b_l \geq 1 - \xi_{ikl}$

$$\xi_{ikl} \geq 0\left(1 \leq i \leq l, (y_k, y_l) \in \hat{Y} \times \acute{Y}\right)$$
$$\sum_{k\in\hat{Y}} P_{ik} \leq \left|\hat{Y}\right| - 1(1 \leq i \leq m, 0 \leq P_{ik} \leq 1)$$
$$P_{ik} = 1\left(k \in \acute{Y}, 1 \leq i \leq m\right)$$

$$\min_{W,\Sigma,P} \sum_{k=1}^{q}\|w_k\|^2 + C_1 \sum_{i=1}^{m} \frac{1}{|\hat{Y}||\acute{Y}|} \sum_{(y_k,y_l)\in\hat{Y}\times\acute{Y}} \widetilde{\boldsymbol{P_{ik}}} * \xi_{ikl}$$

$$-C_2 \sum_{i=1}^{m} \sum_{k\in\hat{Y}_i} S_{k\cdot}(P_{ik}\cdot P_{i\cdot})^T - C_3 \sum_{i=1}^{m} \sum_{k\in\hat{Y}_i} P_{ik}\cdot\|x_i - Q_k\|$$

Subject: $\langle w_k - w_l, x_i\rangle + b_k - b_l \geq 1 - \widetilde{\boldsymbol{P_{ik}}}\xi_{ikl}$

$$\xi_{ikl} \geq 0\left(1 \leq i \leq l, (y_k, y_l) \in \hat{Y} \times \acute{Y}\right)$$

$$\sum_{k\in\hat{Y}} P_{ik} \leq \left|\hat{Y}\right| - 1(1 \leq i \leq m, 0 \leq P_{ik} \leq 1)$$

$$P_{ik} = 1\left(k \in \acute{Y}, 1 \leq i \leq m\right)$$