NeurIPS | 2019
Neural Information Processing Systems

Multi-marginal Wasserstein GAN

Jiezhang Cao[1*], Langyuan Mo[1*], Yifan Zhang[1], Kui Jia[1], Chunhua Shen[3], Mingkui Tan[1,2 †]
[1]South China University of Technology, [2]Peng Cheng Laboratory, [3]The University of Adelaide

South China University of Technology

Problem Definition

- **Multiple marginal matching (M³) problem** aims at learning mappings to match a source domain to multiple target domains.
- **Application:** M³ problem has attracted great attention in many applications, such as multi-domain image translation.

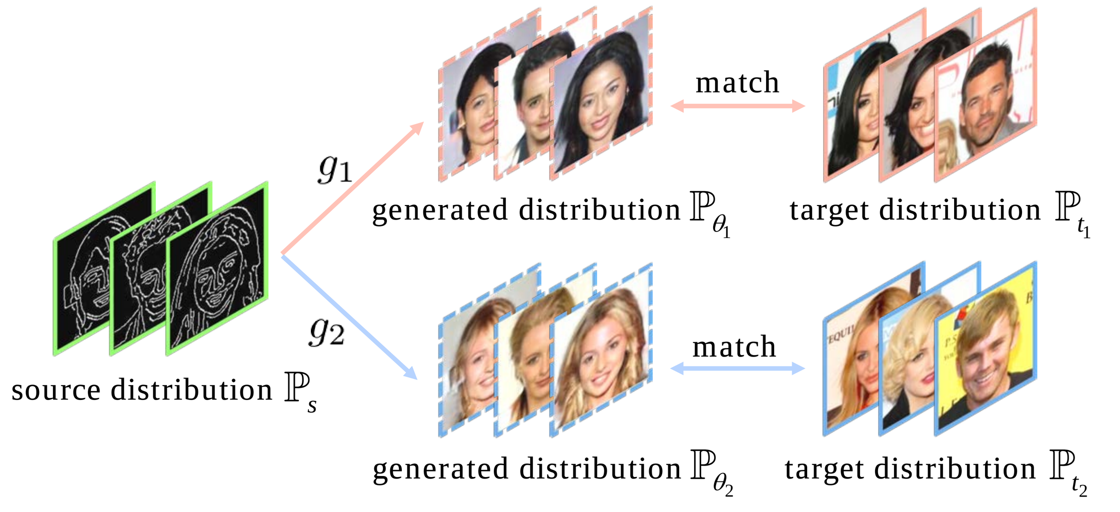


Figure 1: An example of the multi-domain image translation task.

Motivation

- Existing methods neglect to jointly optimize the multi-marginal distance among domains, and thus cannot guarantee the generalization performance and may lead to distribution mismatching issue.
  - StarGAN and UFDN essentially measure the distance between an input distribution and a mixture of all target distributions
- Existing methods ignore correlations among target domains, and are hard to fully capture information to improve the performance.

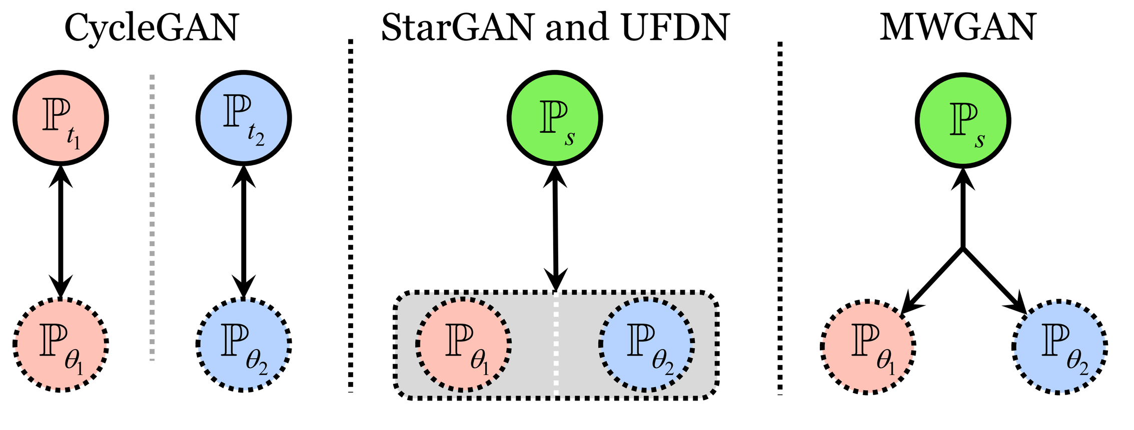


Figure 2: Comparisons of different distribution measures.

Contributions

- We propose a novel MWGAN to optimize the multi-marginal distance among different domains.
- We define and analyze the generalization performance of MWGAN for the multiple domain translation task.
- Extensive experiments demonstrate the effectiveness of MWGAN on balanced and imbalanced translation tasks.

Multi-marginal Wasserstein GAN (MWGAN)

Objective Function. Given a discriminator $f \in \mathcal{F}$ and generators $g_i \in \mathcal{G}$, we define the multi-marginal Wasserstein distance as

$$W\left(\hat{\mathbb{P}}_s, \hat{\mathbb{P}}_{\theta_1}, \ldots, \hat{\mathbb{P}}_{\theta_N}\right) = \max_f \mathbb{E}_{\mathbf{x}\sim\hat{\mathbb{P}}_s}\left[f(\mathbf{x})\right] - \sum\nolimits_i \lambda_i^+ \mathbb{E}_{\hat{\mathbf{x}}\sim\hat{\mathbb{P}}_{\theta_i}}\left[f(\hat{\mathbf{x}})\right],$$

s.t. $\hat{\mathbb{P}}_{\theta_i} \in \mathcal{D}_i, f \in \Omega$, where $\hat{\mathbb{P}}_s$ is the real source distribution, and the distribution $\hat{\mathbb{P}}_{\theta_i}$ is generated by g_i in the i-th domain.

- **Inner-domain constraints:** the generated distribution should belong to the corresponding domain.
- **Inter-domain constraints:** the discriminator should satisfy the inequality constraints w.r.t.the cost function.

Theorem 1 (Generalization bound) *Given the continuous real and generated distributions $\mathbb{P}_s$ and $\mathbb{P}_{\theta_i}, i \in \mathcal{I}$, and the empirical versions $\hat{\mathbb{P}}_s$ and $\hat{\mathbb{P}}_{\theta_i}, i \in \mathcal{I}$ with at least n samples in each domain, there is a universal constant C such that $n \geq C\kappa\Delta^2 \log(L\kappa/\epsilon)/\epsilon^2$ with the error ϵ, the generalization bound is satisfied with probability ($\geq 1-e^{-\kappa}$),*

$$\left|W\left(\hat{\mathbb{P}}_s, \hat{\mathbb{P}}_{\theta_1}, \ldots, \hat{\mathbb{P}}_{\theta_N}\right) - W\left(\mathbb{P}_s, \mathbb{P}_{\theta_1}, \ldots, \mathbb{P}_{\theta_N}\right)\right| \leq \epsilon.$$

- Existing generalization analysis methods study only on two domains and non-trivial for multiple domains.
- MWGAN has a good generalization ability with enough training data when minimizing the multi-domain Wasserstein distance.

Algorithm of MWGAN

Algorithm 1 Multi-marginal WGAN.

Input: Training data $\{\mathbf{x}_j\}_{j=1}^{n_0}$ in the initial domain, $\{\hat{\mathbf{x}}_j^{(i)}\}_{j=1}^{n_i}$ in the i-th target domain; batch size m_{bs}; the number of iterations of the discriminator per generator iteration n_{critic}; Uniform distribution $U[0,1]$

Output: The discriminator f, the generators $\{g_i\}_{i\in[N]}$ and the classifier ϕ.

while not converged **do**
  for $t = 0, \ldots, n_{\text{critic}}$ **do**
    Sample $\mathbf{x}\sim\hat{\mathbb{P}}_s$ and $\hat{\mathbf{x}}\sim\hat{\mathbb{P}}_{\theta_i}, \forall i$, and $\tilde{\mathbf{x}}\leftarrow\rho\mathbf{x}+(1-\rho)\hat{\mathbf{x}}$, $\rho\sim U[0,1]$
    Update discriminator f by ascending the gradient:
$$\nabla_w\left[\mathbb{E}_{\mathbf{x}\sim\hat{\mathbb{P}}_s}\left[f(\mathbf{x})\right] - \sum\nolimits_i \lambda_i^+ \mathbb{E}_{\hat{\mathbf{x}}\sim\hat{\mathbb{P}}_{\theta_i}}\left[f(\hat{\mathbf{x}})\right] + \mathcal{R}_\tau(f)\right]$$
    Update classifier ϕ by descending the gradient $\nabla_v \mathcal{C}_\alpha(\phi)$
  end for
  Update each generator g_i by descending the gradient:
$$\nabla_{\theta_i}\left[-\lambda_i^+ \mathbb{E}_{\hat{\mathbf{x}}\sim\hat{\mathbb{P}}_{\theta_i}}\left[f(\hat{\mathbf{x}})\right] - \mathcal{M}_\alpha(g_i)\right]$$
end while

Balanced Image Translation Results

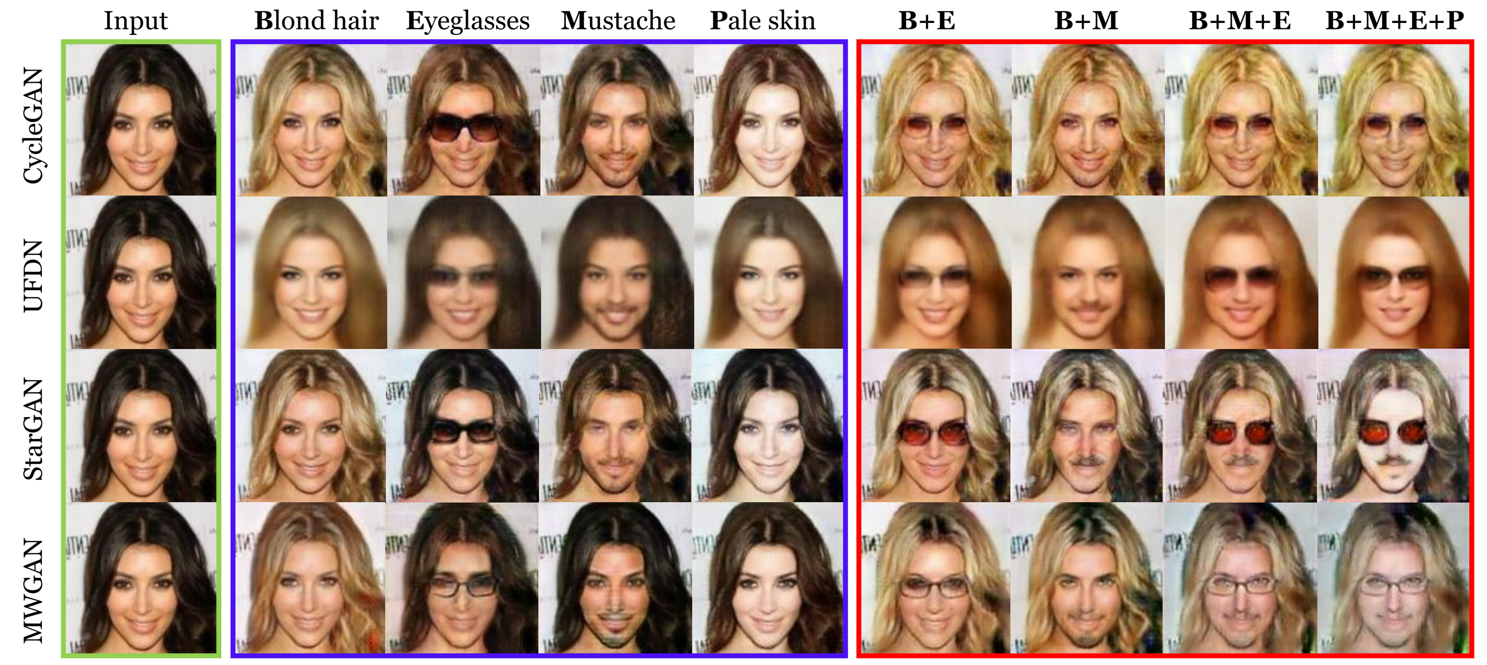


Figure 3: Comparisons of attribute translation on CelebA

- **Single attribute translation task:** MWGAN has a better or comparable performance than baseline methods.
- **Multi-attributes translation task:** MWGAN achieves the highest visual quality, and has good generalization performance.

Imbalanced Image Translation Results

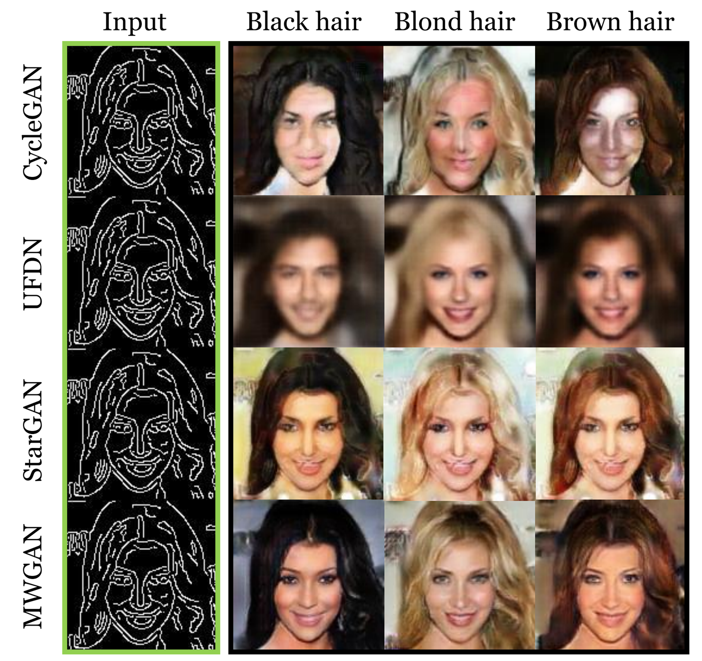


Figure 4: Edge→CelebA translation results

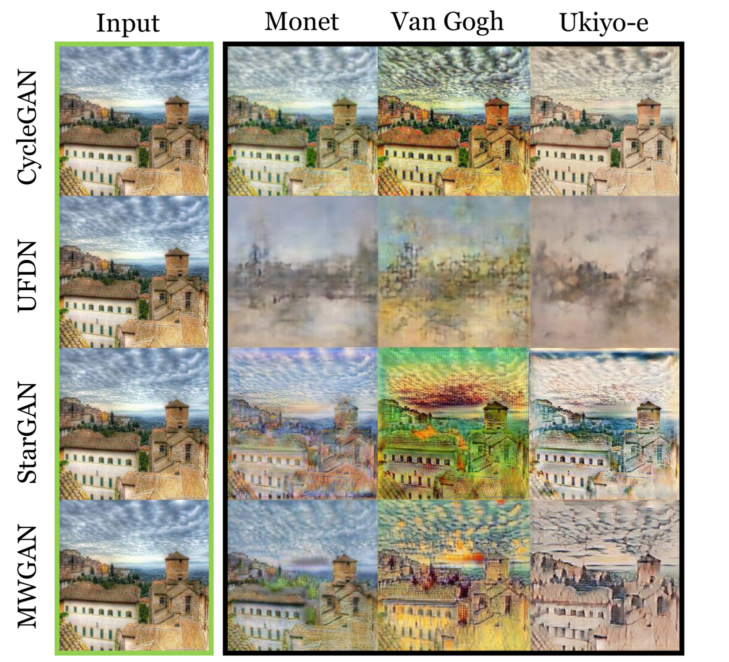


Figure 5: Style transfer results

- **Edge→CelebA task:** MWGAN generates the most natural-looking facial images with the corresponding attributes from edge images.
- **Style transfer task:** MWGAN generates painting images with higher visual quality.

Quantitative Results

Table 1: Comparisons of FID and classification accuracy (%) on single facial attribute translation.

Method	Hair		Eyeglass		Mustache		Pale skin	
	FID	Accuracy (%)	FID	Accuracy (%)	FID	Accuracy (%)	FID	Accuracy (%)
CycleGAN	20.45	95.07	23.69	96.94	24.94	93.89	18.09	80.75
UFDN	65.06	92.01	69.30	79.34	76.04	97.18	53.11	83.33
StarGAN	23.47	96.00	25.36	99.51	23.75	**99.06**	18.12	92.48
MWGAN	**19.63**	**97.65**	**22.94**	**99.53**	**23.69**	98.35	**15.91**	**93.66**

Table 2: Comparisons of classification accuracy (%) on multi-attribute synthesis. (B: Blond hair, E: Eyeglasses, M: Mustache, P: Pale skin.)

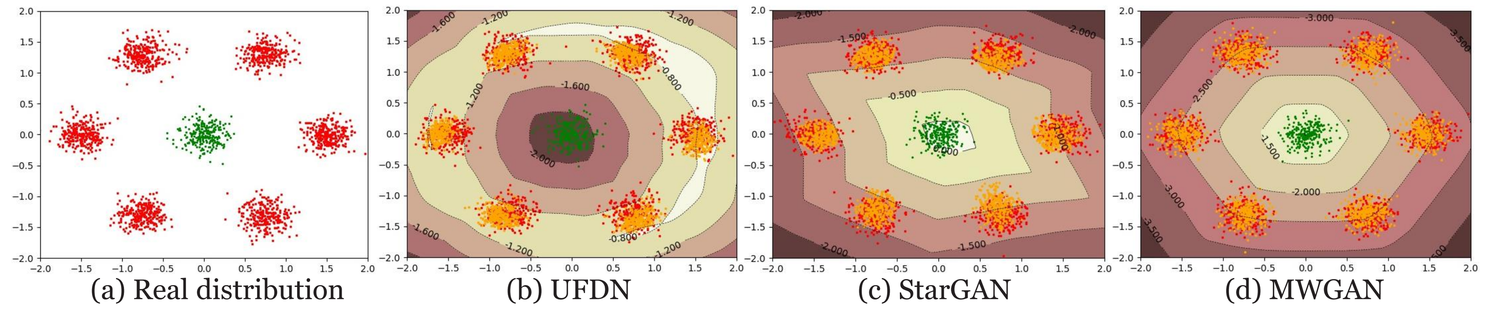
Method	B+E	B+M	B+M+E	B+M+E+P
CycleGAN	66.43	33.33	11.03	2.11
UFDN	72.53	51.40	23.00	8.54
StarGAN	66.66	62.20	45.77	6.10
MWGAN	**75.82**	**69.01**	**53.75**	**19.95**

Table 3: Comparisons of the FID value for each facial attribute (different colors of hair) on the Edge→CelebA translation task.

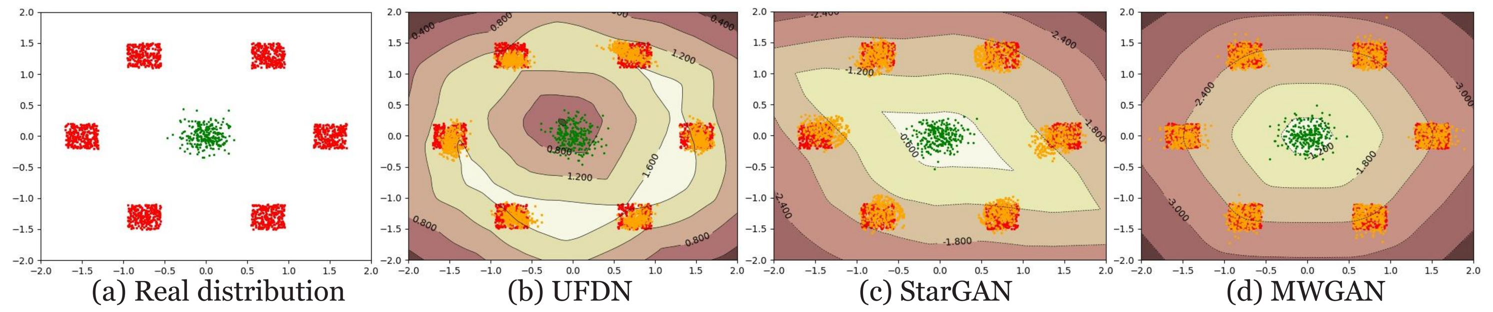
Method	Black hair	Blond hair	Brown hair
CycleGAN	65.10	81.59	65.79
UFDN	131.65	144.78	88.40
StarGAN	53.41	81.00	57.51
MWGAN	**33.81**	**51.87**	**35.24**

- In Tables 1 and 3, MWGAN achieves the lowest FID and comparable accuracy and thus produces the realistic single-attribute.
- In Table 2, MWGAN achieves the highest accuracy and thus synthesizes the most realistic multi-attribute images.

Toy Experiment



(a) Real distribution (b) UFDN (c) StarGAN (d) MWGAN



(a) Real distribution (b) UFDN (c) StarGAN (d) MWGAN

- MWGAN matches the target domain distributions very well, since its discriminator provides correct gradients to update the generators.

Contact Information

- **Name:** Jiezhang Cao
- **Email:** secaojiezhang@mail.scut.edu.cn
- **School:** South China University of Technology
- **Code:** https://github.com/caojiezhang

WeChat