We thank all reviewers for the recognition on the **novelty** and **quality** of our paper: "this work is theoretically motivated, unlike previous works", "state-of-the-art results with high quality ..."(**R2**), "very interesting to discuss the generalization" (**R1**), and "makes a moderate advance for M^3 problem"(**R3**). We first answer a general concern from reviewers.

General Response:

G1. Why use a shared potential function? We address the concern with both **Empirical** and **Theoretical** evidences.

(1) Motivation and empirical justification. We use the shared potential function to exploit the cross-domain correlations for M^3 problem. From Table 6 in Appendix J, more domains indeed help to improve the performance.

(2) Theoretical justification. It is valid to use a shared potential function to replace N ones. In fact, we can prove that the optimal objective of Problem II (with N potential functions) is close to Problem III (with a shared potential function) under mild conditions over $\{\lambda_i\}$ and the cost function. In an extreme case, if $N{+}1$ domains have overlapped samples, the optimal objectives of Problems II and III are equal. These verify the **assumption** in Theorem 1.

Proof sketch: We define the cost function as $c(\mathbf{x}^{(0)}, \ldots, \mathbf{x}^{(N)}) {=} \sum_{i\neq j} d(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$, where $d(\cdot,\cdot)$ is a distance function of two samples and $\mathbf{x}^{(i)}$ is a sample in the i-th domain. The proof can be adapted from the proof of Theorem 3.3 in [24]. For $N{=}1$ (*i.e.*, two domains), if $d(\cdot,\cdot)$ satisfies the triangle inequality, the optimal objective of Problems II and III are equal [24]. Similarly, for $N{\geq}2$, the equivalence holds when $d(\cdot,\cdot)$ satisfies the triangle inequality and $\sum_i \lambda_i{=}0$. Let f^* be an optimizer of Problem III. We first prove $\lambda_0 f^*(\mathbf{x}^{(0)}){=}\inf\{c(\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}) {-} \sum_{j\in[N]} \lambda_j f^*(\mathbf{x}^{(j)})\}$. Let f_i^* be optimal solutions to Problem II and $(f^c)^*$ be the c-conjugate function. If $N{+}1$ domains have overlapped samples $\mathbf{x}_{k_i}^{(i)}{\in}\mathcal{X}^{(i)}$ (see [24]), we can prove $(f^c)^*(\mathbf{x}_{k_0}^{(0)}){=}f_i^*(\mathbf{x}_{k_i}^{(i)}), i{\in}[N]$. Last, we prove any optimal solution to Problem II (resp. III) is a feasible solution to Problem III (resp. II). Then, we conclude the optimal objectives of Problems II and III are equal. For more general cases, we instead prove that the optimal objectives of Problems II and III can be arbitrarily close when multiple domains are very close to each other. We leave the complete proofs in the revised paper. □

To Reviewer #1 (R1):

Q1. More explanations of a shared potential function & Can it exploit correlations? See **General Response G1**.

Q2. Differences of MWGAN from WGAN [3]. MWGAN essentially differs from WGAN even when $\lambda_i^+ = 1/N$: **1)** MWGAN considers and incorporates multi-domain correlations into the inequality constraints to improve the **image translation** performance. WGAN focuses on **image generation tasks** and cannot directly deal with multi-domain correlations. **2)** The objectives of two methods are different in the formulation. **3)** In the algorithm, MWGAN uses gradient penalty to deal with inequality constraints; while WGAN relies on weight clipping.

Q3. Generalization on unseen test samples. Our definition on generalization has considered testing samples (which is similar to [30]). Specifically, in Definition 1, $\mathbb{P}_s$ denotes the probability distribution of unseen source samples.

To Reviewer #2 (R2):

Q1. Theoretical justification of approximation on the potential function. Please refer to **General Response G1**.

Q2. More evaluations with Amazon Mechanical Turk (AMT). We conduct a perceptual evaluation using AMT to assess the performance on the Edge→CelebA translation task, following the settings of StarGAN [6]. From Table A, MWGAN wins significant majority votes for the best perceptual realism, quality and transferred attributes for all facial attributes.

Table A: AMT perceptual evaluation for each attribute.

Method	Black hair	Blond hair	Brown hair
CycleGAN	9.7%	5.7%	9.0%
UFDN	13.2%	15.8%	12.9%
StarGAN	16.0%	21.9%	19.4%
MWGAN	**61.1%**	**56.6%**	**58.7%**

Q3. Order of compositions. We generate attributes with order {Blond hair, Eyeglasses, Mustache and Pale skin}, which works well. The order has a slight impact on the performance. We will include relevant results and discussions.

To Reviewer #3 (R3):

Q1. Empirical and Theoretical sufficiency of a shared potential function. Please refer to **General Response G1**.

Q2. Metrics of domain similarity and its relation to conditions of the shared potential function. The domain similarity/correlation indeed is very critical for our method and theoretical analysis. We start to measure the distance among multiple domains with multi-marginal Wasserstein distance, which however is hard to compute. We thus propose a new feasible dual formulation. From **General Response G1**, if domains are close enough upon sample distances $d(\cdot,\cdot)$, we can use a shared potential function. Nevertheless, in many real problems (*e.g.*, the image translation task), different domains indeed have high correlations, where our method achieved promising performance (See Table A).

Q3. How to understand Fig. 2? Fig. 2 is to show the distribution matching abilities of various methods. The value surface, which depicts the output of the discriminator, is widely used in [14, 24]. More discussions will be included.