Supplementary Material of "ODCR: Orthogonal Decoupling Contrastive Regularization for Unpaired Image Dehazing"

Anonymous CVPR submission

Paper ID 909

A. Implementation Details

We follow the setting of CUT [4], except that the PatchNCE loss is repalced by our Weighted PatchNCE (WPNCE). In detail, we use a 9-block Resnet-based generator [2] with PatchGAN [1]. We define our encoder as the first half of the generator, and accordingly extract our multilayer features from five evenly distributed points of the encoder. For the O-MLPs, we set their numbers of layers to 2 with 256 units in each layer and τ to 0.07.

We train our full model for 400 epochs, keeping the learning rate at $2e^{-4}$ for the first 200 epochs, and linearly decay the learning rate in the last 200 epochs. Each subnet of the network is optimized by the Adam optimizer [3] ($\beta_1 = 0.5, \beta_2 = 0.999$). The samples are randomly cropped into patches of size 256×256 during training.

B. Proofs of Theorems

Theorem 1. Given $\nabla f(\cdot)$ in Euclidean space and $grad\ f(\cdot)$ defined as follows:

$$grad\ f(\Theta) = \nabla f(\Theta) - \frac{1}{2}\Theta\Theta^T\nabla f(\Theta) - \frac{1}{2}\Theta\nabla f(\Theta)^T\Theta \tag{1}$$

$grad\ f(\Theta)$ is the orthogonal projection of $\nabla f(\Theta)$ onto the tangent space of the Stiefel manifold.

Proof 1. If $grad\ f(\cdot)$ is the orthogonal projection of $\nabla f(\cdot)$ onto the tangent space of the Stiefel manifold, according to the triangle law of vector addition, we have:

$$\langle (grad\ f(\cdot) - \nabla f(\cdot)), grad\ f(\cdot)\rangle = 0 \tag{2}$$

where $\langle\cdot,\cdot\rangle$ is the inner product. In particular, for Z_1, Z_2 on the Stiefel manifold:

$$\langle Z_1, Z_2\rangle = tr(Z_1^T Z_2^T) \tag{3}$$

By substituting Eq. 1 and 3 into the left-hand side of Eq. 2, we obtain:

$$\begin{aligned}
&\langle (grad\ f(\Theta) - \nabla f(\Theta)), grad\ f(\Theta)\rangle \\
=&tr(\frac{1}{2}\nabla f(\Theta)^T\Theta\Theta^T\nabla f(\Theta) - \frac{1}{4}\nabla f(\Theta)^T\Theta\Theta^T\Theta\Theta^T\nabla f(\Theta) \\
&- \frac{1}{4}\nabla f(\Theta)^T\Theta\Theta^T\Theta\nabla f(\Theta)^T\Theta + \frac{1}{2}\Theta^T\nabla f(\Theta)\Theta^T\nabla f(\Theta) \\
&- \frac{1}{4}\Theta^T\nabla f(\Theta)\Theta^T\Theta\Theta^T\nabla f(\Theta) - \frac{1}{4}\Theta^T\nabla f(\Theta)\Theta^T\Theta\nabla f(\Theta)^T\Theta)
\end{aligned} \tag{4}$$

Due to $\Theta \in St(p,n)$, Θ satisfies $\Theta^T\Theta = I$. Then Eq. 4 can be simplify as:

$$\begin{aligned}
&\langle (grad\ f(\Theta) - \nabla f(\Theta)), grad\ f(\Theta)\rangle \\
=&\frac{1}{4}tr(\nabla f(\Theta)^T\Theta\Theta^T\nabla f(\Theta)) - \frac{1}{4}tr(\nabla f(\Theta)^T\Theta\nabla f(\Theta)^T\Theta) \\
&+ \frac{1}{4}tr(\Theta^T\nabla f(\Theta)\Theta^T\nabla f(\Theta)) - \frac{1}{4}tr(\Theta^T\nabla f(\Theta)\nabla f(\Theta)^T\Theta)
\end{aligned} \tag{5}$$

Due to $tr(AB) = tr(BA)$ and $tr(A) = tr(A^T)$, we have

$$tr(\nabla f(\Theta)^T\Theta\Theta^T\nabla f(\Theta)) = tr(\Theta^T\nabla f(\Theta)\nabla f(\Theta)^T\Theta) \tag{6}$$

and

$$tr(\Theta^T\nabla f(\Theta)\Theta^T\nabla f(\Theta)) = tr(\nabla f(\Theta)^T\Theta\nabla f(\Theta)^T\Theta) \tag{7}$$

So Eq. 5 equals to 0, demonstrating $grad\ f(\Theta)$ is the orthogonal projection of $\nabla f(\Theta)$. Furthermore, it can also prove that $\Theta^T grad\ f(\Theta) + grad\ f(\Theta)^T f(\Theta) = 0$, which means that $grad\ f(\Theta)$ is on the tangent space.

Theorem 2. Given a point $\Xi = \Theta - \gamma grad\ f(\Theta)$ on the tangent space $T_\Theta St(p,n)$ of a point on a Stiefel manifold, let retraction operation be $\Re_\Theta(\Xi) = (\Theta + \Xi)(I + \Xi^T\Xi)^{-\frac{1}{2}}$, the point Ξ after retraction operation satisfies orthogonal constraint:

$$\Re_\Theta(\Xi)^T\Re_\Theta(\Xi) = I \tag{8}$$

i.e., $\Re_\Theta(\Xi)$ is on $St(p,n)$.

Proof 2. The parameter after retraction $\Re_\Theta$ satisfies:

$$\Re_\Theta(\Xi)^T\Re_\Theta(\Xi) = (I + \Xi^T\Xi)^{-\frac{1}{2}}((\Theta + \Xi)^T(\Theta + \Xi))(I + \Xi^T\Xi)^{-\frac{1}{2}} \tag{9}$$

For Ξ on the tangent space $T_\Theta St(p,n)$, Ξ satisfies:

$$\Theta^T\Xi + \Xi^T\Theta = 0 \tag{10}$$

So

$$\mathfrak{R}_{\Theta}(\Xi)^T\mathfrak{R}_{\Theta}(\Xi) = (I+\Xi^T\Xi)^{-\frac{1}{2}}(I+\Xi^T\Xi)(I+\Xi^T\Xi)^{-\frac{1}{2}} \tag{11}$$

where $(I+\Xi^T\Xi) \in \mathbb{R}^{p\times p}$. Eigenvalue decomposition of $(I+\Xi^T\Xi)$ can get:

$$(I+\Xi^T\Xi) = P\Lambda P^{-1} \tag{12}$$

where Λ stands for a diagonal matrix, and $P^TP = I$. Subsequently, we have:

$$\begin{aligned} \mathfrak{R}_{\Theta}(\Xi)^T\mathfrak{R}_{\Theta}(\Xi) &= P\Lambda^{-\frac{1}{2}}P^TP\Lambda P^TP\Lambda^{-\frac{1}{2}}P^T \\ &= P\Lambda^{-\frac{1}{2}}\Lambda\Lambda^{-\frac{1}{2}}P^T \\ &= PP^T \\ &= I \end{aligned} \tag{13}$$

Thus $\mathfrak{R}_{\Theta}(\Xi)$ satisfies the orthogonal constraint.

References

[1] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017. 1

[2] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, pages 694–711. Springer, 2016. 1

[3] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 1

[4] Taesung Park, Alexei A Efros, Richard Zhang, and Jun-Yan Zhu. Contrastive learning for unpaired image-to-image translation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16*, pages 319–345. Springer, 2020. 1