1. Proof of Proposition 1

Proposition 1 (Jaccard Metric Loss on a hypercube in $\mathbb{R}^D$)**.** *Δ_{JML} is a semi-metric in $[\alpha, \beta]^D \subseteq \mathbb{R}^D$. Specifically, $\forall \mathbf{a}, \mathbf{b} \in [\alpha, \beta]^D$, we have*

(i) Reflexivity: $\Delta_{\text{JML}}(\mathbf{a}, \mathbf{b}) = 0 \iff \mathbf{a} \equiv \mathbf{b}$

(ii) Positivity: $\Delta_{\text{JML}}(\mathbf{a}, \mathbf{b}) \geq 0$

(iii) Symmetry: $\Delta_{\text{JML}}(\mathbf{a}, \mathbf{b}) = \Delta_{\text{JML}}(\mathbf{b}, \mathbf{a})$

Proof. For any $\mathbf{a}, \mathbf{b} \in [\alpha, \beta]^D$, JML is defined in [36] as

$$\Delta_{\text{JML}}(\mathbf{a}, \mathbf{b}) = 1 - \frac{\|\mathbf{a}+\mathbf{b}\|_1 - \|\mathbf{a}-\mathbf{b}\|_1}{\|\mathbf{a}+\mathbf{b}\|_1 + \|\mathbf{a}-\mathbf{b}\|_1}. \tag{1}$$

(i) Reflexivity. If $\Delta_{\text{JML}}(\mathbf{a}, \mathbf{b}) = 0$, we can derive $\|\mathbf{a} - \mathbf{b}\|_1 = \sum_{i=1}^{D} |\mathbf{a}_i - \mathbf{b}_i| = 0$. Thus, we have $\mathbf{a}_i = \mathbf{b}_i, \forall i = 1..D$, which is equivalent to $\mathbf{a} \equiv \mathbf{b}$.

If $\mathbf{a} \equiv \mathbf{b}$, we obviously have $\Delta_{\text{JML}}(\mathbf{a}, \mathbf{b}) = 0$.

(ii) Positivity. The property is satisfied because we can rewrite Δ_{JML} as follows

$$\Delta_{\text{JML}}(\mathbf{a}, \mathbf{b}) = \frac{2\,\|\mathbf{a}-\mathbf{b}\|_1}{\|\mathbf{a}+\mathbf{b}\|_1 + \|\mathbf{a}-\mathbf{b}\|_1} \geq 0, \forall \mathbf{a}, \mathbf{b} \in [\alpha, \beta]^D \tag{2}$$

(iii) Symmetry. As $\|\mathbf{a} + \mathbf{b}\|_1 = \|\mathbf{b} + \mathbf{a}\|_1$ and $\|\mathbf{a} - \mathbf{b}\|_1 = \|\mathbf{b} - \mathbf{a}\|_1, \forall \mathbf{a}, \mathbf{b} \in [\alpha, \beta]^D$, we obviously have $\Delta_{\text{JML}}(\mathbf{a}, \mathbf{b}) = \Delta_{\text{JML}}(\mathbf{b}, \mathbf{a})$ and this concludes the proof. □

2. Proof of Proposition 2

Lemma 2.1. *Let $\ell : [-1, 1] \times [-1, 1] \to \mathbb{R}$ be defined by $\ell(x, y) = |y - x||y - x|^{\gamma \mathbb{I}(yx \geq 0)}$. For any fixed $y_0 \in [0, 1]$ (or $[-1, 0]$), the function $\ell(x, y_0)$ does not have any local infimum at $x \in [-1, 0)$ (or $(0, 1]$).*

Proof. As $\ell(x, y_0)$ is symmetric, without loss of generality, we assume that $y_0 \in [0, 1]$. For simplicity, we denote $\ell(x) = \ell(x, y_0), \forall x \in [-1, 1]$. First, we rewrite $\ell(x)$ as

$$\ell(x) = \begin{cases} |x - y_0|^{\gamma+1} & xy_0 \geq 0 \\ |x - y_0| & \text{otherwise} \end{cases} \tag{3}$$

$\forall x \in [-1, 0)$, the function ℓ becomes a decreasing linear function

$$\ell(x) = y_0 - x \tag{4}$$

Therefore, the only potential local infimum is at $x \to 0^-$. However, we have that

$$\lim_{x \to 0^-} \ell(x) = y_0 \tag{5}$$

$$\geq y_0^{\gamma+1} \qquad \triangleright \text{ For } \gamma \geq 1 \text{ and } y_0 \in [0, 1] \tag{6}$$

$$= \lim_{x \to 0^+} \ell(x) \qquad \triangleright\ y_0 > 0 \tag{7}$$

If $y_0 \neq 0$, then $\lim_{x \to 0^-} \ell(x) > \lim_{x \to 0^+} \ell(x)$. Thus, $x \to 0^-$ is not a local infimum. On the other hand, if $y_0 = 0$, then $x = y_0 = 0 \notin [-1, 0)$. Here, we conclude the proof. □

Proposition 2 (Stable Focal-L1)**.** *Let $\ell : [-1, 1] \times [-1, 1] \to \mathbb{R}$ be defined by $\ell(x, y) = |y - x||y - x|^{\gamma \mathbb{I}(yx \geq 0)}$. Given an arbitrary fixed $y_0 \in [-1, 1]$, we have that $\ell(x, y_0)$ has only one strictly local and global minimum at $x = y_0$.*

Proof. Let $\ell : [-1, 1] \times [-1, 1] \to \mathbb{R}$ be defined by $\ell(x, y) = |y - x||y - x|^{\gamma \mathbb{I}(yx \geq 0)}$. Consider an arbitrary fixed $y_0 \in [-1, 1]$. As $\ell(x, y_0)$ is symmetric, without loss of generality, we assume that $y_0 \in [0, 1]$. For simplicity, we denote $\ell(x) = \ell(x, y_0), \forall x \in [-1, 1]$. First, we rewrite $\ell(x)$ as

$$\ell(x) = \begin{cases} |x - y_0|^{\gamma+1} & xy_0 \geq 0 \\ |x - y_0| & \text{otherwise} \end{cases} \tag{8}$$

(i) $y_0 \in (0, 1]$**:** $\forall x \in [0, 1]$, we have that

$$\ell(x) = |x - y_0|^{\gamma+1} \tag{9}$$

One can observe that

$$\ell(x) > \ell(y_0) = 0, \forall x \in [0, 1] \backslash \{y_0\} \tag{10}$$

Consider an arbitrary $x \in [-1, 0)$, ℓ then becomes a decreasing linear function, that is

$$\ell(x) = y_0 - x \tag{11}$$

Then, we have the following derivations: $\forall x \in [-1, 0)$,

$$\ell(x) \geq \inf_{x \in [-1, 0)} \ell(x) \tag{12}$$

$$= \lim_{x \to 0^-} \ell(x) \tag{13}$$

$$= y_0 \qquad \triangleright \text{ For (11)} \tag{14}$$

$$> y_0^{\gamma+1} \qquad \triangleright \text{ For } \gamma \geq 1 \text{ and } y_0 \in (0, 1] \tag{15}$$

$$= \ell(0) \qquad \triangleright \text{ For (9) and } y_0 > 0 \tag{16}$$

$$> \ell(y_0) = 0 \qquad \triangleright \text{ For (10)} \tag{17}$$

From (10) and (17), we can infer that

$$\ell(x) > \ell(y_0), \forall x \in [-1, 1] \backslash \{y_0\}. \tag{18}$$

In other words, $x = y_0$ is the only strictly global minimum of ℓ in $[-1, 1]$.

(ii) $y_0 = 0$**:** We have that

$$\ell(x) = |x|^{\gamma+1} \tag{19}$$

Similarly, one can observe that

$$\ell(0) < \ell(x), \forall x \in [-1,1] \backslash \{0\}, \tag{20}$$

which implies that $x = 0$ is the only strictly global minimum in $[-1, 1]$.

From (i) and (ii), we conclude that $x = y_0$ is the only strictly global minimum of $\ell(x)$ in $[-1, 1]$.

Furthermore, $\ell(x)$ is a convex function in $[0, 1]$ as its second derivative is non-negative in this domain, that is

$$\frac{\partial^2}{\partial x^2}\ell(x) = 2(\gamma+1)\delta(x-y_0)|x-y_0|^{\gamma} \tag{21}$$

$$+ \gamma(\gamma+1)(x-y_0)^2|x-y_0|^{\gamma-3} \geq 0, \forall x \in [0,1] \tag{22}$$

where δ is the Dirac Delta function. Thus, ℓ has at most one local minimum in $[0, 1]$, which is $x = y_0$. Together with Lemma 2.1, we conclude that the function $\ell(x)$ has only one strictly local and global minimum at $x = y_0$ in $[-1, 1]$. □

3. Proof of Proposition 3

Proposition 3 (Stable Focal-L1 as a lower bound of Focal-L1)**.** $\mathcal{L}^S_{\text{FocalL1}}(\mathbf{f}, \tilde{\mathbf{y}}) \leq \mathcal{L}_{\text{FocalL1}}(\mathbf{f}, \tilde{\mathbf{y}}), \forall \mathbf{f}, \tilde{\mathbf{y}} \in [-1, 1]^D$.

Proof. We need to prove that $\mathcal{L}^S_{\text{FocalL1}}(\mathbf{f}, \tilde{\mathbf{y}}) \leq \mathcal{L}_{\text{FocalL1}}(\mathbf{f}, \tilde{\mathbf{y}}), \forall \mathbf{f}, \tilde{\mathbf{y}} \in [-1, 1]^D$.

We denote that

$$\ell^S(\tilde{\mathbf{y}}_i, \mathbf{f}_i) = |\tilde{\mathbf{y}}_i - \mathbf{f}_i||\tilde{\mathbf{y}}_i - \mathbf{f}_i|^{\gamma \mathbb{I}(\tilde{\mathbf{y}}_i \mathbf{f}_i \geq 0)}, \tag{23}$$

$$\ell(\tilde{\mathbf{y}}_i, \mathbf{f}_i) = |\tilde{\mathbf{y}}_i - \mathbf{f}_i| \frac{|\tilde{\mathbf{y}}_i - \mathbf{f}_i|^{\gamma \mathbb{I}(\tilde{\mathbf{y}}_i \mathbf{f}_i \geq 0)}}{\max(|\tilde{\mathbf{y}}_i|, |\mathbf{f}_i|)}. \tag{24}$$

Then, the two losses become

$$\mathcal{L}^S_{\text{FocalL1}}(\tilde{\mathbf{y}}, \mathbf{f}) = \frac{1}{|\Omega|}\sum_{i \in \Omega} \ell^S(\tilde{\mathbf{y}}_i, \mathbf{f}_i), \tag{25}$$

$$\mathcal{L}_{\text{FocalL1}}(\tilde{\mathbf{y}}, \mathbf{f}) = \frac{1}{|\Omega|}\sum_{i \in \Omega} \ell(\tilde{\mathbf{y}}_i, \mathbf{f}_i). \tag{26}$$

Because $\max(|\tilde{\mathbf{y}}_i|, |\mathbf{f}_i|) \leq 1, \forall \tilde{\mathbf{y}}_i, \mathbf{f}_i \in [-1, 1]$, we straightforwardly derive that $\ell^S(\tilde{\mathbf{y}}_i, \mathbf{f}_i) \leq \ell(\tilde{\mathbf{y}}_i, \mathbf{f}_i), \forall i \in \Omega$. Thus, we can conclude the proof. □

4. Image Processing Details

In the LR setting, the images were resized to a fixed size of 512×512. These images were augmented by random cropping and random zooming, as well as other augmentations such as random flipping, rotation, color jittering, gamma correction, Gaussian noises, and cutout. We employed the Lanczos interpolation filter during resizing, with a radius of 3, and anti-aliasing scaling. During inference, the same resizing procedure was applied.

In the HR setting, the training patches were generated in "offline" and "online" fashions. Regarding the offline generation, we cropped the patches via a sliding window, where the window size was in turn 256×256, 512×512, 512×768, 768×1024, and 1024×1024, before using scale augmentation. Meanwhile, the online generation created randomly cropped patches from the images, where the size of the cropped patches ranged from 20% to 70% of the original size. The aspect ratio could vary from 0.8 to 1.2. Both the offline and online patches were augmented for training. Similar to the LR setting, we used the Lanczos interpolation filter during resizing, with a radius of 3 and anti-aliasing scaling. During inference, we used a sliding window to get patches from the image, and these patches were resized to 512×512 using the same filter. The window size and the step size were empirically chosen among several values to achieve the best Dice score on each dataset. The window size was 256 for DRIVE and STARE, 512 for CHASEDB1, and 1024 for FIVES and HRF. The step size ranged from 128 for DRIVE and STARE to 256 for other.

Table S1. Throughput and FLOPs results

Method	Images/s	FLOPs
IterNet [17]	0.67	2367.2G
FR-UNet [19]	0.34	10872.8G
DUNet [32]	0.24	648.0G
CE-Net [11]	0.59	1408.4G
UNet++ [39]	1.27	11293.7G
UNet [25]	3.37	2104.4G
CTF-Net [35]	1.74	1370.0G
MAGF-Net [16]	0.20	22253.3G
UNet [25]	38.89	42.9G
UNet++ [39]	30.45	230.4G
Swin-UNet [6]	49.52	55.3G
D2SF [23]	15.66	353.8G
DA-Net [33]	37.31	174.1G
Teacher-Student	30.45	230.4G
GeoLS [31, 36]	30.45	230.4G
LS [30, 36]	30.45	230.4G
BLS [36]	30.45	230.4G
SiNGR [8]	30.45	230.4G
Ours (DA-Net)	35.90	174.1G
Ours (Swin-UNet)	49.52	55.3G
Ours (UNet)	41.67	42.9G
Ours (UNet++)	30.45	230.4G