D-Convexity: A Unified Differentiable Convex Shape Prior via Quasi-Concavity for Data-driven Image Segmentation

Supplementary Material

A. Proof of Theorem 1

Zero-order quasi-concavity condition: $u \in C^0$ is quasi-concave $\Longleftrightarrow$ For any $\mathbf{x}, \mathbf{y} \in \Omega$, $\lambda \in [0,1]$, $u(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \geq \min\{u(\mathbf{x}), u(\mathbf{y})\}$.

Proof. ($\Rightarrow$) Denote $\gamma = \min\{u(\mathbf{x}), u(\mathbf{y})\}$. Apparently $u(\mathbf{x}), u(\mathbf{y}) \geq \gamma$, which means $\mathbf{x}, \mathbf{y} \in S_\gamma$. Given that u is quasi-concave, its super-lever set S_γ is convex. Therefore, any convex combination of $\mathbf{x}, \mathbf{y}$, i.e. $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in S_\gamma$. This means $u(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \geq \gamma = \min\{u(\mathbf{x}), u(\mathbf{y})\}$.

($\Leftarrow$) For any γ, consider $\mathbf{x}, \mathbf{y} \in S_\gamma$. By definition, $u(\mathbf{x}), u(\mathbf{y}) \geq \gamma$. Therefore, $u(\lambda\mathbf{x} + (1-\lambda)\mathbf{y}) \geq \min\{u(\mathbf{x}), u(\mathbf{y})\} \geq \gamma$. This means the convex combination $\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in S_\gamma$, i.e. u's super-level set is convex, and u is quasi-concave. □

B. Proof of Lemma 1

Supporting hyperplane given by the gradient: Let $u : \Omega \to \mathbb{R}$ be a C^1 function. Fix $\gamma \in \mathbb{R}$ and consider the super-level set $S_\gamma := \{\mathbf{x} \in \Omega | u(\mathbf{x}) \geq \gamma\}$. Assume S_γ is convex. Let $\mathbf{y} \in \partial S_\gamma$ be a boundary point with $u(\mathbf{y}) = \gamma$ and suppose $\nabla u(\mathbf{y}) \neq 0$. Then the affine hyperplane,

$$T_\mathbf{y} := \{\mathbf{x} \in \Omega | \nabla u(\mathbf{y})^\top (\mathbf{x} - \mathbf{y}) = 0\} \tag{15}$$

is a supporting hyperplane of S_γ at $\mathbf{y}$, i.e. S_γ is contained in the closed half-space

$$S_\gamma \subset H_\mathbf{y} := \{\mathbf{x} \in \Omega | \nabla u(\mathbf{y})^\top (\mathbf{x} - \mathbf{y}) \geq 0\}. \tag{16}$$

In particular, $T_\mathbf{y}$ is also tangent to the contour $\partial S_\gamma = \{\mathbf{x} \in \Omega | u(\mathbf{x}) = \gamma\}$, therefore the normal vector of any supporting hyperplane at $\mathbf{y}$ can be chosen parallel to $\nabla u(\mathbf{y})$.

Proof. Since $u \in C^1$, the first-order Taylor expansion at $\mathbf{y}$ gives, for $\mathbf{x}$ near $\mathbf{y}$,

$$\nu(\mathbf{x}) := u(\mathbf{x}) - u(\mathbf{y}) = \nabla u(\mathbf{y})^\top (\mathbf{x} - \mathbf{y}) + r(\mathbf{x}), \tag{17}$$

$$\text{where } \frac{r(\mathbf{x})}{\|\mathbf{x} - \mathbf{y}\|} \to 0 \text{ as } \mathbf{x} \to \mathbf{y}. \tag{18}$$

Take any $\mathbf{x} \in S_\gamma$ (so $u(\mathbf{x}) \geq u(\mathbf{y})$). Because S_γ is convex and $\mathbf{y} \in \partial S_\gamma$, the line segment

$$\ell(t) := \mathbf{y} + t(\mathbf{x} - \mathbf{y}),\ t \in [0,1], \tag{19}$$

lies in S_γ. Define the scalar function $\phi(t) := u(\ell(t))$. Note that $\phi(0) = u(\mathbf{y})$ and $\phi(1) = u(\mathbf{x}) \geq u(\mathbf{y})$, hence $\phi(t) \geq u(\mathbf{y})$ for all $t \in [0,1]$ because each point of the segment belongs to S_γ. Differentiating ϕ at $t = 0$ yields

$$\phi'(0) = \nabla u(\mathbf{y})^\top (\mathbf{x} - \mathbf{y}). \tag{20}$$

Since $\phi(t) \geq \phi(0)$ for small positive t, we must have $\phi'(0) \geq 0$, i.e.

$$\nabla u(\mathbf{y})^\top (\mathbf{x} - \mathbf{y}) \geq 0. \tag{21}$$

This inequality exactly says that the affine hyperplane $T_\mathbf{y}$ is supporting, since every $\mathbf{x} \in S_\gamma$ lies in the closed half-space determined by $T_\mathbf{y}$.

Since $\nabla u(\mathbf{y}) \neq 0$, the Implicit Function Theorem implies that the contour $\{\mathbf{x} \in \Omega | u(\mathbf{x}) = \gamma\}$ is a C^1 hypersurface in a neighborhood of $\mathbf{y}$. Its tangent space at $\mathbf{y}$ is the kernel of $\nabla u(\mathbf{y})$, i.e.

$$\begin{aligned} T_\mathbf{y} &= \ker(\nabla u(\mathbf{y})) = \{\mathbf{d} | \ \nabla u(\mathbf{y})^\top \mathbf{d} = 0\} \\ &= \{\mathbf{x} \in \Omega | \nabla u(\mathbf{y})^\top (\mathbf{x} - \mathbf{y}) = 0\}. \end{aligned} \tag{22}$$

If a supporting hyperplane at $\mathbf{y}$ has normal $\mathbf{n}$, then by uniqueness of the tangent (since u is differentiable and $\nabla u(\mathbf{y}) \neq 0$) we have $\mathbf{n}$ parallel to $\nabla u(\mathbf{y})$, which proves the final claim. □

C. Proof of Theorem 2

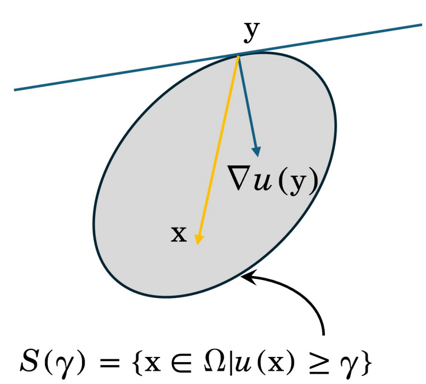


Figure 5. Illustration of the first-order condition. If $u(\mathbf{x}) \geq u(\mathbf{y})$, then vectors $\mathbf{x} - \mathbf{y}$ and $\nabla u(\mathbf{y})$ must form an acute angle.

First-order quasi-concavity condition: $u \in C^1$ is quasi-concave $\Longleftrightarrow$ If $u(\mathbf{x}) \geq u(\mathbf{y})$, then $\nabla u(\mathbf{y})^\top (\mathbf{x} - \mathbf{y}) \geq 0$.

Proof. ($\Rightarrow$) Suppose u is quasi-concave. Then every super-level set $S_\gamma = \{\mathbf{x} \in \Omega | u(\mathbf{x}) \geq \gamma\}$ is convex. According to Lemma 1, for any boundary point $\mathbf{y}$ with $u(\mathbf{y}) = \gamma$, if $\nabla u(\mathbf{y}) = 0$, the condition holds trivially. Otherwise, S_γ is contained in the closed half space of $T_{\mathbf{y}}$, i.e.

$$\begin{aligned} &u(\mathbf{x}) \geq \gamma = u(\mathbf{y}) \Rightarrow \mathbf{x} \in S_\gamma \\ &\Rightarrow \mathbf{x} \in H_{\mathbf{y}} = \{\mathbf{x} \in \Omega | \nabla u(\mathbf{y})^\top (\mathbf{x} - \mathbf{y}) \geq 0\}. \end{aligned} \tag{23}$$

Since γ is arbitrary, for any $\mathbf{x}, \mathbf{y}$ with $u(\mathbf{x}) \geq u(\mathbf{y})$, we have $\nabla u(\mathbf{y})^\top (\mathbf{x} - \mathbf{y}) \geq 0$.

($\Leftarrow$) Conversely, fix any γ and consider the super-level set S_γ. Take arbitrary $\mathbf{x}_0, \mathbf{x}_1 \in S_\gamma$ and define

$$\phi(t) = u((1-t)\mathbf{x}_0 + t\mathbf{x}_1),\ t \in [0, 1]. \tag{24}$$

If ϕ attained a value strictly below $\min\{\phi(0), \phi(1)\}$ at an interior $t_0 \in (0, 1)$, w.l.o.g, let $\phi(1) \geq \phi(0) > \phi(t_0)$. According to continuity and Intermediate Value Theorem, there exists an interval $[t_0, t_1] \subset [0, 1]$ such that $\phi(t_0) < \phi(t_1) = \phi(0)$, and $\phi(t) \leq \phi(0)$ when $t \in [t_0, t_1]$. Such an interval exists by continuity, and can be found easily by extending $(0, \phi(0))$ along t axis and check the intersections with $\phi(t)$.

Take $t \in (t_0, t_1)$, and let $\mathbf{x}_t = (1-t)\mathbf{x}_0 + t\mathbf{x}_1$. By the hypothesis, since both $u(\mathbf{x}_0), u(\mathbf{x}_1) \geq u(\mathbf{x}_t)$,

$$\nabla u(\mathbf{x}_t)^\top (\mathbf{x}_i - \mathbf{x}_t) \geq 0,\ i = 0, 1, \tag{25}$$

which gives

$$\begin{aligned} t\nabla u(\mathbf{x}_t)^\top (\mathbf{x}_0 - \mathbf{x}_1) \geq 0, \\ (1-t)\nabla u(\mathbf{x}_t)^\top (\mathbf{x}_0 - \mathbf{x}_1) \leq 0, \end{aligned} \tag{26}$$

collectively,

$$\nabla u(\mathbf{x}_t)^\top (\mathbf{x}_1 - \mathbf{x}_0) = 0,\ \forall t \in (t_0, t_1). \tag{27}$$

According to Mean Value Theorem, there exists $t^* \in (t_0, t_1)$, such that

$$\begin{aligned} 0 < \phi(t_1) - \phi(t_0) &= \nabla u(\mathbf{x}_{t^*})^\top (\mathbf{x}_{t_1} - \mathbf{x}_{t_0}) \\ &= (t_1 - t_0)\nabla u(\mathbf{x}_{t^*})^\top (\mathbf{x}_1 - \mathbf{x}_0) = 0, \end{aligned} \tag{28}$$

which is contradictory. Therefore $\phi(t) \geq \min\{\phi(0), \phi(1)\}$ for all t, so the segment $[\mathbf{x}_0, \mathbf{x}_1] \subset S_\gamma$. Hence S_γ is convex. Since γ was arbitrary, u is quasi-concave. □

D. Proof of Theorem 5

Second-order quasi-concavity sufficient condition: Let $u \in C^2$, if the Hessian $\nabla^2 u(\mathbf{x}) \prec 0$ (strict negative definite) on tangent space $T_{\mathbf{x}}$ for all $\mathbf{x} \in \Omega$ such that $\nabla u(\mathbf{x}) \neq 0$, then $u \in C^2$ is quasi-concave.

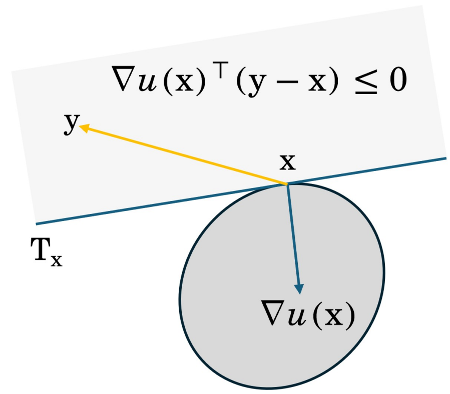


Figure 6. Illustration of the second-order condition. For fixed $\mathbf{x}$, if u is quasi-concave, then on the half space $\nabla u(\mathbf{x})^\top (\mathbf{y} - \mathbf{x}) \leq 0$, maximum value of u is taken at $\mathbf{x}$.

Proof. Fix any γ and consider the super-level set S_γ. Take arbitrary $\mathbf{x}_0, \mathbf{x}_1 \in S_\gamma$ and define

$$\begin{aligned} &\phi(t) = u((1-t)\mathbf{x}_0 + t\mathbf{x}_1),\ t \in [0, 1], \\ &\mathbf{x}_t = (1-t)\mathbf{x}_0 + t\mathbf{x}_1. \end{aligned} \tag{29}$$

If ϕ attained a value strictly below $\min\{\phi(0), \phi(1)\}$ at an interior $t_0 \in (0, 1)$, without loss of generality, let $\phi(1) \geq \phi(0) > \phi(t_0)$. Then according to continuity and Extreme Value Theorem, ϕ attains a strict interior minimum at some $t^* \in (0, 1)$ such that $\phi(1) \geq \phi(0) > \phi(t^*)$. At this t^*, we have $\phi'(t^*) = 0, \phi''(t^*) \geq 0$. Further, $\phi'(t^*) = 0$ implies

$$\phi'(t^*) = \nabla u(\mathbf{x}_{t^*})^\top (\mathbf{x}_1 - \mathbf{x}_0) = 0, \tag{30}$$

which means $\mathbf{x}_1 - \mathbf{x}_0 \in T_{\mathbf{x}_{t^*}}$. According to the negative definite hypothesis,

$$(\mathbf{x}_1 - \mathbf{x}_0)^\top \nabla^2 u(\mathbf{x}_{t^*})(\mathbf{x}_1 - \mathbf{x}_0) < 0. \tag{31}$$

However, this contradicts with

$$\phi''(t^*) = (\mathbf{x}_1 - \mathbf{x}_0)^\top \nabla^2 u(\mathbf{x}_{t^*})(\mathbf{x}_1 - \mathbf{x}_0) \geq 0. \tag{32}$$

Therefore, for all t, $\phi(t) \geq \min\{\phi(0), \phi(1)\}$, so the segment $[\mathbf{x}_0, \mathbf{x}_1] \subset S_\gamma$. Hence S_γ is convex. Since γ was arbitrary, u is quasi-concave. □

E. Explicit form of convex loss gradient

In the proposed convex gradient projection module, the term $\nabla_u \mathcal{L}_{convex}(u)$ can be either computed by auto-gradient, or by using an explicit form, and both approaches are efficient. Let $D_x, D_y, D_{xx}, D_{yy}, D_{xy}$ be fixed linear convolution operators with adjoints $D_x^\top, D_y^\top, D_{xx}^\top, D_{yy}^\top, D_{xy}^\top$ realizing first and second order finite differences. Specifically,

$$D_x = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 1 & 0 \end{bmatrix},\ D_{xx} = D_x^\top D_x = \begin{bmatrix} 0 & 1 & 0 \\ 0 & -2 & 0 \\ 0 & 1 & 0 \end{bmatrix},$$
$$D_y = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{bmatrix}, D_{yy} = D_y^\top D_y = \begin{bmatrix} 0 & 0 & 0 \\ 1 & -2 & 1 \\ 0 & 0 & 0 \end{bmatrix},$$
$$D_{xy} = D_{yx} = \frac{1}{2}(D_y D_x + D_x D_y) = \frac{1}{2}\begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 1 & -1 \end{bmatrix}. \tag{33}$$

For any function $u : \Omega \to [0, 1]$ define

$$\begin{aligned} u_x &= D_x u, \quad u_y = D_y u, \\ u_{xx} &= D_{xx} u, \quad u_{yy} = D_{yy} u, \quad u_{xy} = D_{xy} u. \end{aligned} \tag{34}$$

Explicit gradient of $\mathcal{L}_{2nd}$. The loss is

$$\mathcal{L}_{2nd}(u) = \frac{1}{|\Omega|} \sum_{\mathbf{x} \in \Omega} \|\nabla u(\mathbf{x})\| R(\mathbf{x}), \tag{35}$$

where $R = \mathrm{ReLU}(Q_2 + \delta)$, and $H = \mathbf{1}[Q_2 + \delta > 0]$ (indicator function). Therefore,

$$\begin{aligned} \nabla_u \mathcal{L}_{2nd}(u) = \frac{1}{|\Omega|} \sum_{\mathbf{x} \in \Omega} [\nabla_u \|\nabla u(\mathbf{x})\| R(\mathbf{x}) \\ + \|\nabla u(\mathbf{x})\| \nabla_u R(\mathbf{x})]. \end{aligned} \tag{36}$$

With $Q_2 = u_x^2 u_{yy} - 2u_x u_y u_{xy} + u_y^2 u_{xx}$, the pointwise partial derivatives are

$$\begin{aligned} &\frac{\partial Q_2}{\partial u_x} = 2u_x u_{yy} - 2u_y u_{xy},\ \frac{\partial Q_2}{\partial u_y} = 2u_y u_{xx} - 2u_x u_{xy}, \\ &\frac{\partial Q_2}{\partial u_{xx}} = u_y^2,\ \frac{\partial Q_2}{\partial u_{yy}} = u_x^2,\ \frac{\partial Q_2}{\partial u_{xy}} = -2u_x u_y. \end{aligned} \tag{37}$$

Since $u_x = D_x u$ etc. are linear in u, the chain rule with operator adjoints gives the discrete gradient

$$\begin{aligned} \nabla_u Q_2(u) = D_{xx}^\top(u_y^2) + D_{yy}^\top(u_x^2) + D_{xy}^\top(-2u_x u_y) \\ + D_x^\top(2u_x u_{yy} - 2u_y u_{xy}) + D_y^\top(2u_y u_{xx} - 2u_x u_{xy}). \end{aligned} \tag{38}$$

For numerical stability use the smoothed magnitude

$$\|\nabla u\| = \sqrt{u_x^2 + u_y^2 + \varepsilon} \quad (\varepsilon > 0), \tag{39}$$

then

$$\frac{\partial \|\nabla u\|}{\partial u_x} = \frac{u_x}{\|\nabla u\|}, \quad \frac{\partial \|\nabla u\|}{\partial u_y} = \frac{u_y}{\|\nabla u\|}, \tag{40}$$

and the chain rule yields

$$\nabla_u \|\nabla u\| = D_x^\top\left(\frac{u_x}{\|\nabla u\|}\right) + D_y^\top\left(\frac{u_y}{\|\nabla u\|}\right). \tag{41}$$

Substituting above with $\nabla_u R = H \nabla_u Q_2$ gives

$$\begin{aligned} \nabla_u \mathcal{L}_{\text{2nd}}(u) = \frac{1}{|\Omega|} \Big[& \underbrace{D_x^\top\Big(R \frac{u_x}{\|\nabla u\|}\Big) + D_y^\top\Big(R \frac{u_y}{\|\nabla u\|}\Big)}_{\text{through } g = \|\nabla u\|_\varepsilon} \\ & + \underbrace{D_x^\top\Big(\|\nabla u\| H\,(2u_x u_{yy} - 2u_y u_{xy})\Big)}_{\text{via } \partial Q_2/\partial u_x} \\ & + \underbrace{D_y^\top\Big(\|\nabla u\| H\,(2u_y u_{xx} - 2u_x u_{xy})\Big)}_{\text{via } \partial Q_2/\partial u_y} \\ & + \underbrace{D_{xx}^\top(\|\nabla u\| H\, u_y^2) + D_{yy}^\top(\|\nabla u\| H\, u_x^2)}_{\text{via } \partial Q_2/\partial u_{xx}, \partial Q_2/\partial u_{yy}} \Big] \\ & + \underbrace{D_{xy}^\top(\|\nabla u\| H\,(-2u_x u_y))}_{\text{via } \partial Q_2/\partial u_{xy}} \Big]. \end{aligned} \tag{42}$$

With the elementwise sigmoid $u = \sigma(o) = 1/(1 + e^{-o})$ one has $\frac{\partial u}{\partial o} = u\,(1 - u)$, hence

$$\nabla_o \mathcal{L}_{\text{2nd}}(u(o)) = \big(\nabla_u \mathcal{L}_{\text{2nd}}(u)\big) \odot u\,(1 - u).$$

Explicit gradient of $\mathcal{L}_{1st}$. Let D_x, D_y be fixed linear convolution operators and write

$$\nabla u(\mathbf{y}) = \big(D_x u(\mathbf{y}),\, D_y u(\mathbf{y})\big)^\top. \tag{43}$$

For a temperature $\varepsilon > 0$ define the smoothed sigmoid

$$\sigma_\varepsilon(t) = \frac{1}{1 + e^{-t/\varepsilon}},\ \sigma'_\varepsilon(t) = \frac{1}{\varepsilon}\sigma_\varepsilon(t)\big(1 - \sigma_\varepsilon(t)\big). \tag{44}$$

Given a neighborhood $\mathcal{N}_\mathbf{y} \subset \Omega$ around each $\mathbf{y}$, set for every ordered pair $(\mathbf{x}, \mathbf{y})$ with $\mathbf{x} \in \mathcal{N}_\mathbf{y}$:

$$\begin{aligned} &S(\mathbf{x}, \mathbf{y}) = \sigma_\varepsilon\big(u(\mathbf{x}) - u(\mathbf{y})\big), \quad \mathbf{v}_{\mathbf{x},\mathbf{y}} = \mathbf{x} - \mathbf{y}, \\ &R(\mathbf{x}, \mathbf{y}) = H(\mathbf{x}, \mathbf{y})\,\big(-\nabla u(\mathbf{y})^\top \mathbf{v}_{\mathbf{x},\mathbf{y}}\big), \end{aligned} \tag{45}$$

where $H(\mathbf{x}, \mathbf{y}) = \mathbf{1}[\nabla u(\mathbf{y})^\top \mathbf{v}_{\mathbf{x},\mathbf{y}} < 0]$. The loss reads

$$\mathcal{L}_{\text{1st}}(u) = \frac{1}{|\Omega|} \sum_{\mathbf{y} \in \Omega} \sum_{\mathbf{x} \in \mathcal{N}_\mathbf{y}} S(\mathbf{x}, \mathbf{y})\, R(\mathbf{x}, \mathbf{y}).$$

For any pixel $p \in \Omega$, the partial derivative $\big[\nabla_u \mathcal{L}_{1st}(u)\big](p)$ collects all terms in which $u(p)$ appears. There are three kinds of appearances.

(A) $u(p)$ as a "source" value $u(\mathbf{x})$ in $S(\mathbf{x}, \mathbf{y})$. Here $u(\mathbf{x})$ contributes when $\mathbf{x} = p$ and $p \in \mathcal{N}_\mathbf{y}$:

$$\frac{\partial S(\mathbf{x}, \mathbf{y})}{\partial u(p)} = \begin{cases} \sigma'_\varepsilon\big(u(p) - u(\mathbf{y})\big), & \mathbf{x} = p, \\ 0, & \mathbf{x} \neq p. \end{cases} \tag{46}$$

Thus the contribution is

$$\sum_{\mathbf{y}:\, p\in\mathcal{N}_{\mathbf{y}}} \sigma'_\varepsilon\big(u(p)-u(\mathbf{y})\big)\, R(p,\mathbf{y}). \tag{47}$$

(B) $u(p)$ **as an "anchor" value** $u(\mathbf{y})$ **in** $S(\mathbf{x},\mathbf{y})$**.** Here $u(\mathbf{y})$ contributes when $\mathbf{y}=p$ and $\mathbf{x}\in\mathcal{N}_p$:

$$\frac{\partial S(\mathbf{x},\mathbf{y})}{\partial u(p)} = \begin{cases} -\sigma'_\varepsilon\big(u(\mathbf{x})-u(p)\big), & \mathbf{y}=p, \\ 0, & \mathbf{y}\neq p. \end{cases} \tag{48}$$

Thus the contribution is

$$-\sum_{\mathbf{x}\in\mathcal{N}_p} \sigma'_\varepsilon\big(u(\mathbf{x})-u(p)\big)\, R(\mathbf{x},p). \tag{49}$$

(C) $u(p)$ **through the anchor gradient** $\nabla u(\mathbf{y})$ **inside** $R(\mathbf{x},\mathbf{y})$**.** $R(\mathbf{x},\mathbf{y})$ depends on u through the anchor gradient $\nabla u(\mathbf{y})$. Collecting all such contributions over $\mathbf{y}$, the resulting gradient can be written compactly using the adjoint operators $D_x^\top, D_y^\top$.

Write $v_x(\mathbf{x},\mathbf{y})=(\mathbf{x}-\mathbf{y})_x$, $v_y(\mathbf{x},\mathbf{y})=(\mathbf{x}-\mathbf{y})_y$. Then

$$R(\mathbf{x},\mathbf{y}) = -H(\mathbf{x},\mathbf{y})\Big(D_x u(\mathbf{y})\, v_x(\mathbf{x},\mathbf{y}) + D_y u(\mathbf{y})\, v_y(\mathbf{x},\mathbf{y})\Big). \tag{50}$$

Introduce two auxiliary fields $C_x, C_y : \Omega\to\mathbb{R}$,

$$\begin{aligned} C_x(p) &= -\sum_{\mathbf{x}\in\mathcal{N}_p} S(\mathbf{x},p)\, H(\mathbf{x},p)\, v_x(\mathbf{x},p), \\ C_y(p) &= -\sum_{\mathbf{x}\in\mathcal{N}_p} S(\mathbf{x},p)\, H(\mathbf{x},p)\, v_y(\mathbf{x},p), \end{aligned} \tag{51}$$

Therefore, the part of the objective depending on R can be rewritten as

$$\begin{aligned} &\sum_{\mathbf{y}\in\Omega}\sum_{\mathbf{x}\in N_{\mathbf{y}}} S(\mathbf{x},\mathbf{y})R(\mathbf{x},\mathbf{y}) \\ &= \sum_{\mathbf{y}\in\Omega} C_x(\mathbf{y})\, D_x u(\mathbf{y}) + \sum_{\mathbf{y}\in\Omega} C_y(\mathbf{y})\, D_y u(\mathbf{y}). \end{aligned} \tag{52}$$

To differentiate it, note that both D_x, D_y are linear operators, and we use the following standard identity: for any linear operator A and any coefficient field c,

$$\sum_{\mathbf{y}} c(\mathbf{y})\,(Au)(\mathbf{y}) = c^\top A u = (A^\top c)^\top u, \tag{53}$$

hence

$$\nabla_u\left(\sum_{\mathbf{y}} c(\mathbf{y})\,(Au)(\mathbf{y})\right) = A^\top c. \tag{54}$$

Applying this identity to the two terms gives

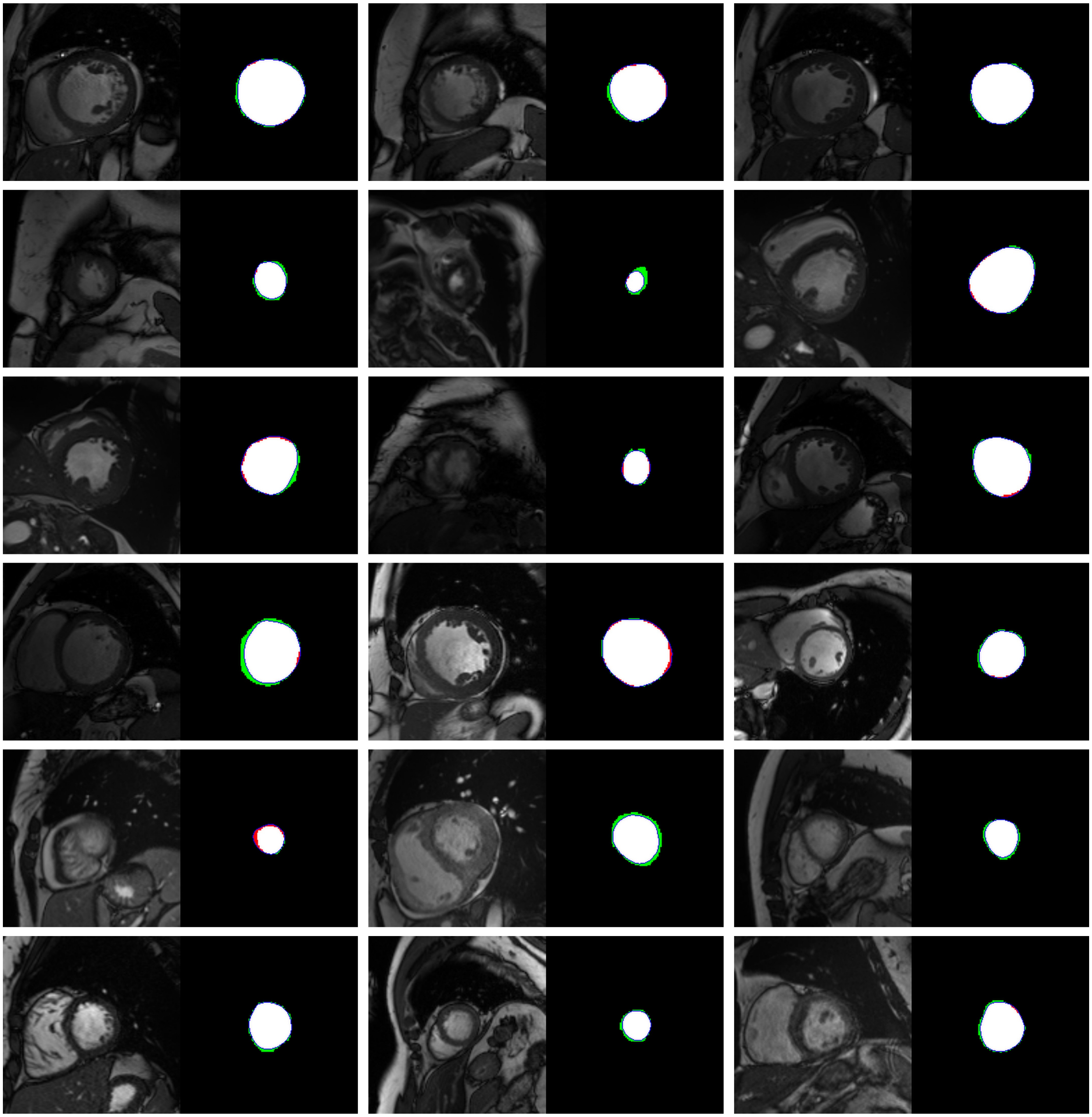
$$\begin{aligned} &\left(\nabla_u \sum_{\mathbf{y}\in\Omega}\sum_{\mathbf{x}\in N_{\mathbf{y}}} S(\mathbf{x},\mathbf{y})R(\mathbf{x},\mathbf{y})\right)_{\text{via } R} \\ &= D_x^\top C_x + D_y^\top C_y. \end{aligned} \tag{55}$$

Summing (A)+(B)+(C),

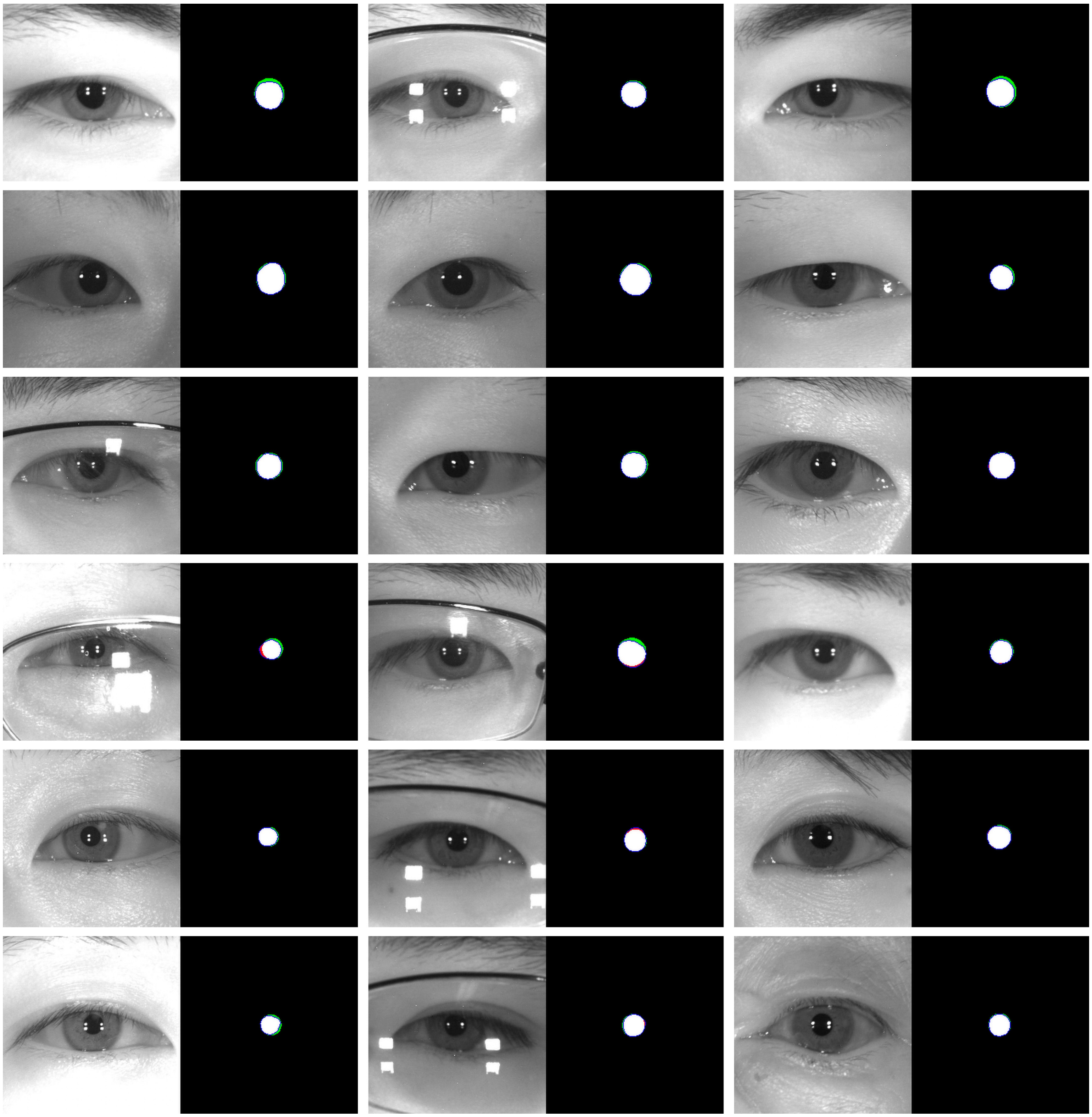
$$\begin{aligned} \big[\nabla_u \mathcal{L}_{1\text{st}}(u)\big](p) = \frac{1}{|\Omega|}\Big[&\underbrace{\sum_{\mathbf{y}:\, p\in\mathcal{N}_{\mathbf{y}}} \sigma'_\varepsilon\big(u(p)-u(\mathbf{y})\big)\, R(p,\mathbf{y})}_{\text{via } \partial S/\partial u} \\ &- \underbrace{\sum_{\mathbf{x}\in\mathcal{N}_p} \sigma'_\varepsilon\big(u(\mathbf{x})-u(p)\big)\, R(\mathbf{x},p)}_{\text{via } \partial S/\partial u} \\ &+ \underbrace{\big[D_x^\top C_x + D_y^\top C_y\big](p)}_{\text{via } \partial R/\partial u}\Big], \end{aligned} \tag{56}$$

If $u=\sigma(o)$ is the elementwise sigmoid of logits o, then $\nabla_o \mathcal{L}_{1\text{st}}(u(o)) = \big(\nabla_u \mathcal{L}_{1\text{st}}(u)\big)\odot u(1-u)$ elementwise.

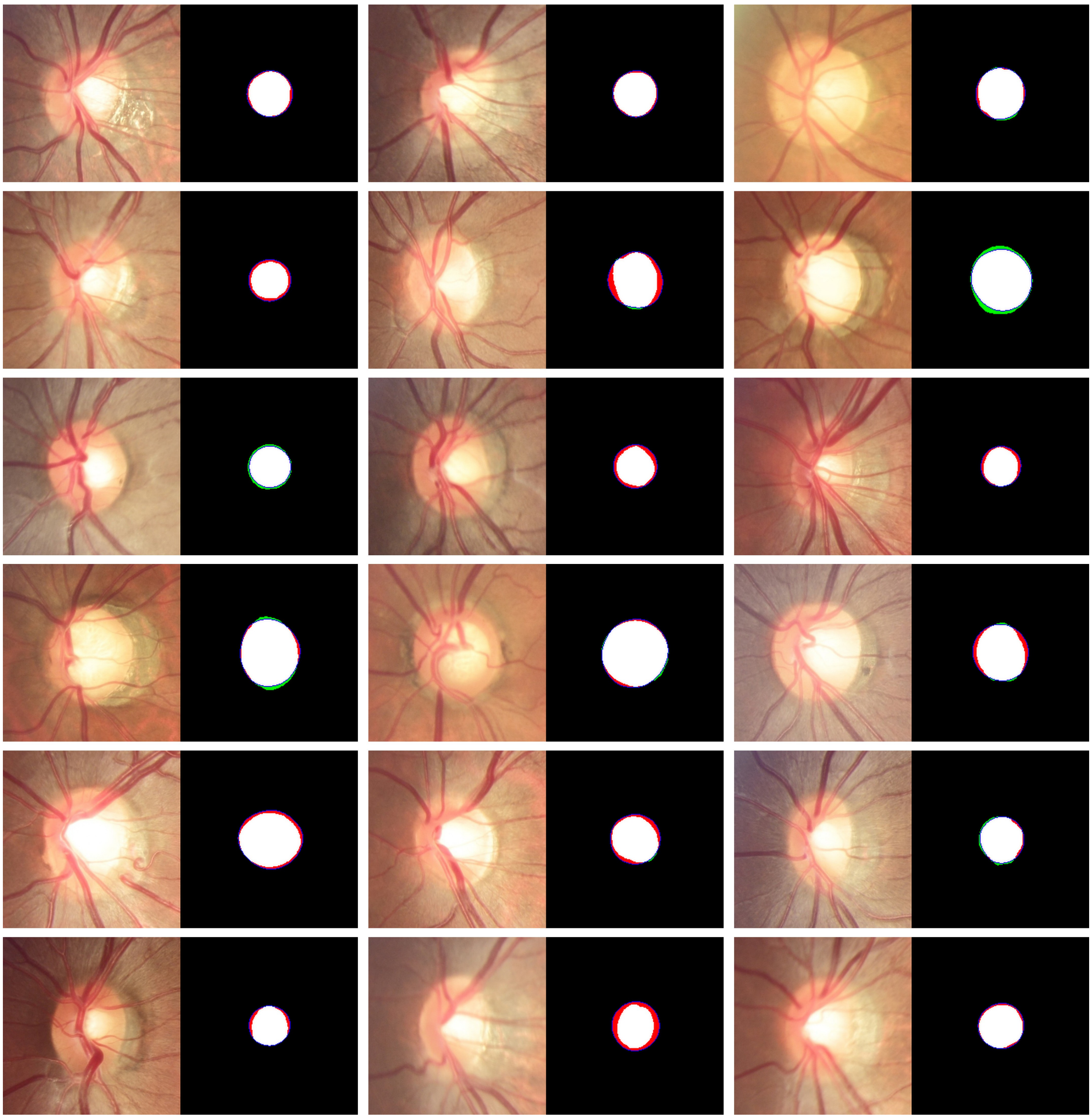
F. Qualitative Visualization on ACDC Dataset.



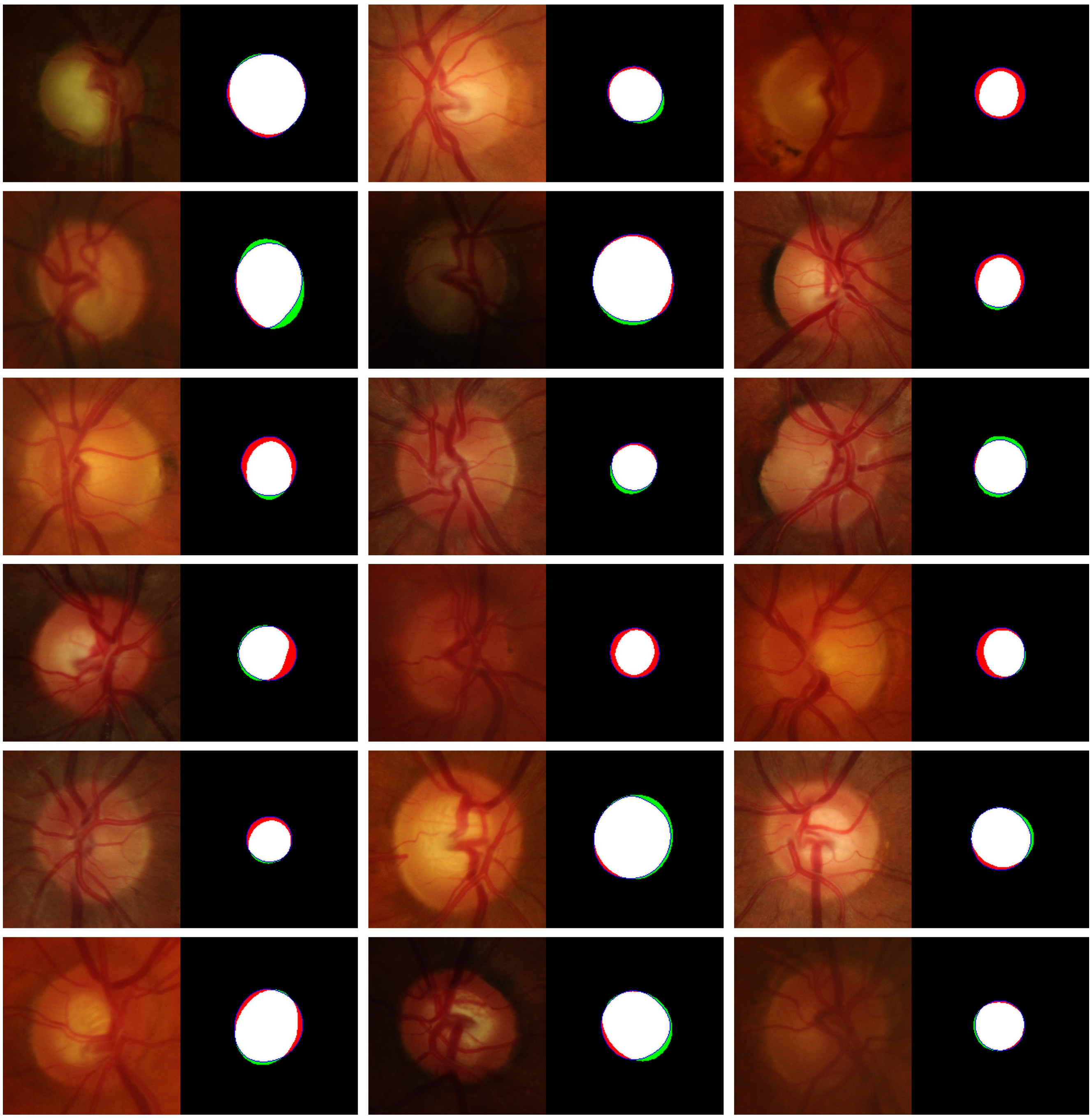
G. Qualitative Visualization on CASIA Dataset.



H. Qualitative Visualization on REFUGE Dataset.



I. Qualitative Visualization on RIM-ONE-r3 Dataset.



J. Ten Trials Results for the Ablation Study

Trial	Prior	Backbone	Dice	IoU	HD	Backbone	Dice	IoU	HD	Backbone	Dice	IoU	HD
0	-	U-Net	83.02	70.97	12.56	Swin-Unet	85.32	74.40	7.530	DeepLabV3p	74.52	59.39	12.18
	$0th$		82.93	70.83	12.52		85.19	74.20	7.516		74.49	59.35	12.17
	$1st$		83.59	71.81	11.67		86.13	75.64	7.093		76.66	62.15	11.09
	$2nd$		84.78	73.58	9.837		88.53	79.43	5.969		82.12	69.67	8.527
1	-	U-Net	84.28	72.83	11.04	Swin-Unet	86.65	76.44	6.965	DeepLabV3p	77.21	62.88	11.04
	$0th$		84.22	72.74	10.99		86.56	76.31	6.942		77.16	62.81	11.04
	$1st$		84.88	73.74	10.31		87.23	77.35	6.464		79.02	65.31	10.10
	$2nd$		86.30	75.90	8.764		88.74	79.75	5.665		82.93	70.84	8.244
2	-	U-Net	83.66	71.91	11.99	Swin-Unet	88.72	79.73	6.137	DeepLabV3p	75.97	61.25	11.42
	$0th$		83.31	71.39	11.86		88.69	79.67	6.117		75.95	61.23	11.42
	$1st$		83.83	72.16	11.02		89.31	80.69	5.783		77.96	63.88	10.42
	$2nd$		84.85	73.69	9.687		90.73	83.03	5.043		82.80	70.65	8.183
3	-	U-Net	85.48	74.65	9.804	Swin-Unet	86.29	75.89	8.377	DeepLabV3p	77.54	63.32	10.74
	$0th$		85.50	74.68	9.640		86.14	75.65	8.256		77.53	63.30	10.73
	$1st$		85.97	75.39	8.973		87.11	77.17	7.542		79.89	66.51	9.618
	$2nd$		86.09	75.58	8.055		88.79	79.84	7.082		84.66	73.41	7.390
4	-	U-Net	84.98	73.88	10.20	Swin-Unet	87.01	77.01	10.59	DeepLabV3p	77.20	62.87	11.26
	$0th$		84.93	73.80	10.16		86.81	76.69	10.50		77.19	62.85	11.25
	$1st$		85.56	74.76	9.380		87.44	77.68	8.022		79.17	65.52	10.31
	$2nd$		86.50	76.21	8.106		87.95	78.49	7.349		83.80	72.12	8.092
5	-	U-Net	83.74	72.02	11.86	Swin-Unet	84.35	72.94	7.777	DeepLabV3p	78.40	64.48	10.93
	$0th$		83.57	71.78	11.78		84.26	72.81	7.777		78.36	64.41	10.93
	$1st$		84.25	72.79	11.02		85.13	74.11	7.330		80.38	67.20	9.616
	$2nd$		85.66	74.92	8.977		87.51	77.80	6.243		84.25	72.78	7.580
6	-	U-Net	83.95	72.34	11.49	Swin-Unet	83.07	71.05	8.197	DeepLabV3p	76.80	62.34	12.20
	$0th$		83.68	71.94	11.40		82.99	70.92	8.190		76.81	62.35	12.20
	$1st$		84.17	72.67	10.68		83.80	72.12	7.773		79.62	66.14	10.18
	$2nd$		85.03	73.96	9.489		85.90	75.28	6.847		85.43	74.57	7.370
7	-	U-Net	83.68	71.94	11.69	Swin-Unet	84.58	73.28	8.912	DeepLabV3p	77.72	63.55	10.72
	$0th$		83.44	71.59	11.58		84.11	72.58	8.783		77.69	63.52	10.71
	$1st$		84.01	72.43	10.91		85.38	74.48	7.938		79.99	66.66	9.525
	$2nd$		85.33	74.41	9.558		87.64	77.99	6.789		83.95	72.34	7.462
8	-	U-Net	85.03	73.96	10.39	Swin-Unet	86.99	76.98	6.555	DeepLabV3p	78.96	65.24	10.39
	$0th$		84.91	73.77	10.30		86.93	76.89	6.539		78.96	65.23	10.38
	$1st$		85.52	74.70	9.608		87.63	77.98	6.169		80.39	67.21	9.661
	$2nd$		86.85	76.76	8.281		89.27	80.62	5.431		83.97	72.37	7.832
9	-	U-Net	83.50	71.67	11.96	Swin-Unet	85.82	75.16	8.329	DeepLabV3p	77.73	63.57	11.19
	$0th$		83.27	71.33	11.77		85.27	74.32	8.193		77.69	63.51	11.18
	$1st$		83.83	72.16	11.09		86.23	75.79	7.387		79.54	66.03	10.22
	$2nd$		85.03	73.96	9.832		87.49	77.76	6.648		83.83	72.16	8.148