Supplementary Materials for CVPR2020 accepted paper:

Computing Valid p-values for Image Segmentation by Selective Inference

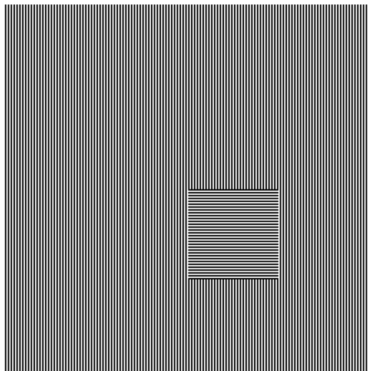
A Specific examples that any test statistic can be used in our framework under certain conditions

The basic assumption of "large object-background mean intensity difference" equals "reliable segmentation mask" may be not valid for some segmentation tasks. For some segmentation tasks, the object and background may have similar or even almost identical intensities. However, any test statistic can be used in our framework if it is represented as a linear combination of pixel intensities, i.e., test statistics after certain filterbank-based representations such as Fourier, Wavelet, Gabor transforms can be also fit into our framework.

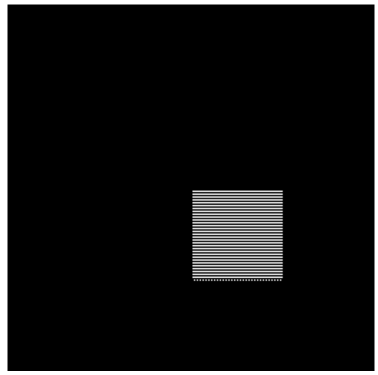
For example, when convolving a certain 3×3 linear filter that components are represented by $\boldsymbol{vec}(F) := [\phi_1, \dots, \phi_9]^\top$, (e.g., a first-order differential filter of the y-axis direction F_1 is represented by $\boldsymbol{vec}(F_1) := [0, 0, 0, -1, 1, 0, 0, 0, 0]^\top$), the image feature after conversion $\boldsymbol{x}'$ is represented by the following.

$$\begin{aligned} x'_i =& \phi_1 x_{i-w-1} + \phi_2 x_{i-w} + \phi_3 x_{i-w+1} + \phi_4 x_{i-1} + \phi_5 x_i \\ &+ \phi_6 x_{i+1} + \phi_7 x_{i+w-1} + \phi_8 x_{i+w} + \phi_9 x_{i+w+1} \end{aligned}$$

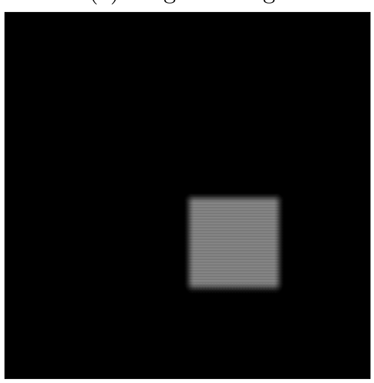
where, w is the width of the image. We can use this $\boldsymbol{x}'$ instead of $\boldsymbol{x}$. However, note that selection event corresponding to $\boldsymbol{x}$ also changed to $\boldsymbol{x}'$. Figure 1 shows an example that statistical tests are performed using the test statistic after filter transformation in our framework. By performing the filtering process, two regions can be detected even if there are no differences in average intensities in the original image. Furthermore, it can be observed that the p-values obtained with the proposed method are smaller than $\alpha = 0.05$.



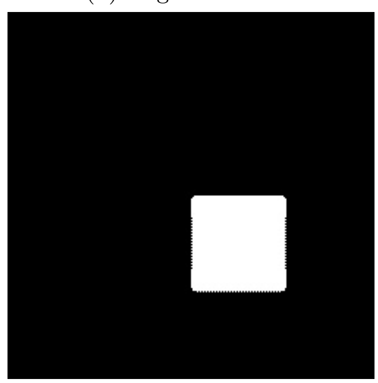
(a) Original image



(b) Edge extraction



(c) Blurred



(d) Segmentation result

Figure 1: An example of performing a statistical test using feature values after filter operation. (a) There are two regions with different textures, but there is no difference in the average intensity of the two areas. (b) The result of applying a first-order differential filter of the y-axis direction to a. (c) The result of applying 5×5 mean filter to b. (d) The result of applying TH-based segmentation. As a result, selective-$p = \mathbf{0.00}$ was obtained.

B Proof of Theorem 1

To formally define a valid p-value, the difference between random variables and corresponding observations must be clarified. In the rest of this section, for a random variable a, $a^{\rm obs}$ represents the corresponding observation. For notational simplicity, let us write the test statistic as $\Delta = |\delta|$ with $\delta = m_{\rm ob} - m_{\rm bj}$. Then, the conditional p-value for the observed difference $\Delta^{\rm obs}$ in Thm. 1 is formally written as

$$\mathfrak{p}(\Delta^{\rm obs}) = \mathbb{P}_{\rm H_0}\left(\Delta \geq \Delta^{\rm obs} \;\middle|\; \begin{array}{c} \mathcal{A}(\boldsymbol{x}) = \mathcal{A}(\boldsymbol{x}^{\rm obs}) \\ \boldsymbol{z}(\boldsymbol{x}) = \boldsymbol{z}(\boldsymbol{x}^{\rm obs}) \\ {\rm sgn}(\delta) = {\rm sgn}(\delta^{\rm obs}) \end{array}\right). \tag{1}$$

By definition, the p-value function $\mathfrak{p}(\cdot)$ in (1) should satisfy

$$\mathbb{P}_{\rm H_0}\left(\mathfrak{p}(\Delta^{\rm obs}) \leq \alpha \;\middle|\; \begin{array}{c} \mathcal{A}(\boldsymbol{x}) = \mathcal{A}(\boldsymbol{x}^{\rm obs}) \\ \boldsymbol{z}(\boldsymbol{x}) = \boldsymbol{z}(\boldsymbol{x}^{\rm obs}) \\ {\rm sgn}(\delta) = {\rm sgn}(\delta^{\rm obs}) \end{array}\right) = \alpha \quad \forall \alpha \in [0,1]. \tag{2}$$

Since the property (2) is satisfied if and only if

$$\left[\mathfrak{p}(\Delta) \;\middle|\; \begin{array}{c} \mathcal{A}(\boldsymbol{x}) = \mathcal{A}(\boldsymbol{x}^{\rm obs}) \\ \boldsymbol{z}(\boldsymbol{x}) = \boldsymbol{z}(\boldsymbol{x}^{\rm obs}) \\ {\rm sgn}(\delta) = {\rm sgn}(\delta^{\rm obs}) \end{array}\right] \sim {\rm Unif}[0,1],$$

we prove the validity of the proposed p-value computation method

$$\mathfrak{p}(\Delta^{\rm obs}) = 1 - F_{0,\boldsymbol{\eta}^\top \Sigma \boldsymbol{\eta}}^{E(\boldsymbol{z}(\boldsymbol{x}^{\rm obs}))}(\Delta^{\rm obs}) \tag{3}$$

by showing that

$$\left[1 - F_{0,\boldsymbol{\eta}^\top \Sigma \boldsymbol{\eta}}^{E(\boldsymbol{z}(\boldsymbol{x}^{\rm obs}))}(\Delta) \;\middle|\; \begin{array}{c} \mathcal{A}(\boldsymbol{x}) = \mathcal{A}(\boldsymbol{x}^{\rm obs}) \\ \boldsymbol{z}(\boldsymbol{x}) = \boldsymbol{z}(\boldsymbol{x}^{\rm obs}) \\ {\rm sgn}(\delta) = {\rm sgn}(\delta^{\rm obs}) \end{array}\right] \sim {\rm Unif}[0,1] \tag{4}$$

in the following proof.

Proof. The difference in the average pixel intensities between the object and background regions is written as

$$\Delta = |\delta| = |m_{\rm ob} - m_{\rm bg}| = \left|\frac{1}{|\mathcal{O}|}\sum_{p \in \mathcal{O}} x_p - \frac{1}{|\mathcal{B}|}\sum_{p \in \mathcal{B}} x_p\right| = \boldsymbol{\eta}^\top \boldsymbol{x}$$

where

$$\boldsymbol{\eta} = {\rm sgn}(\delta)\left(\frac{1}{|\mathcal{O}|}\boldsymbol{e}_{\mathcal{O}} - \frac{1}{|\mathcal{B}|}\boldsymbol{e}_{\mathcal{B}}\right)$$

Consider a decomposition[1] of $\boldsymbol{x}$ into two independent components $\boldsymbol{z}$ and $\boldsymbol{w}$ such that

$$\boldsymbol{x} = \boldsymbol{z}(\boldsymbol{x}) + \boldsymbol{w}, \text{ where } \boldsymbol{z}(\boldsymbol{x}) = (I_n - \frac{\Sigma \boldsymbol{\eta}\boldsymbol{\eta}^\top}{\boldsymbol{\eta}^\top \Sigma \boldsymbol{\eta}})\boldsymbol{x}, \text{ and } \boldsymbol{w} = \frac{\Sigma \boldsymbol{\eta}\boldsymbol{\eta}^\top}{\boldsymbol{\eta}^\top \Sigma \boldsymbol{\eta}}\boldsymbol{x}.$$

[1]In the case of $\Sigma = I_n$, this decomposition indicates the projection of $\boldsymbol{x}$ to $\boldsymbol{\eta}$ and its orthogonal complement.

Since $\boldsymbol{w}$ is written as $\boldsymbol{w} = \Delta \boldsymbol{y}$ with $\boldsymbol{y} = \frac{\Sigma \boldsymbol{\eta}^\top}{\boldsymbol{\eta}^\top \Sigma \boldsymbol{\eta}}$, $\boldsymbol{x}$ is represented as

$$\boldsymbol{x} = \boldsymbol{z}(\boldsymbol{x}) + \Delta \boldsymbol{y}.$$

Thus, by fixing $\boldsymbol{z}(\boldsymbol{x}) = \boldsymbol{z}(\boldsymbol{x}^{\text{obs}})$ by conditioning, the quadratic inequality conditions

$$\boldsymbol{x}^\top A_j \boldsymbol{x} + \boldsymbol{b}_j^\top \boldsymbol{x} + c_j \le 0, j = 1, 2, \ldots$$

specify the range of the test statistic as $\Delta \in E(\boldsymbol{z}(\boldsymbol{x}^{\text{obs}}))$ with

$$\begin{aligned} &E(\boldsymbol{z}(\boldsymbol{x}^{\text{obs}})) \\ &= \bigcap_j \{\Delta \ge 0 \mid (\boldsymbol{z}(\boldsymbol{x}^{\text{obs}}) + \Delta \boldsymbol{y})^\top A_j (\boldsymbol{z}(\boldsymbol{x}^{\text{obs}}) + \Delta \boldsymbol{y}) + \boldsymbol{b}_j^\top (\boldsymbol{z}(\boldsymbol{x}^{\text{obs}}) + \Delta \boldsymbol{y}) + c_j \le 0\}. \end{aligned}$$

Under the conditions that $\mathcal{A}(\boldsymbol{x}) = \mathcal{A}(\boldsymbol{x}^{\text{obs}})$, $\boldsymbol{z}(\boldsymbol{x}) = \boldsymbol{z}(\boldsymbol{x}^{\text{obs}})$ and $\text{sgn}(\delta) = \text{sgn}(\delta^{\text{obs}})$, $\boldsymbol{\eta}$ is considered as a non-random fixed vector, and since $\boldsymbol{x}$ is normally distributed, $\Delta = \boldsymbol{\eta}^\top \boldsymbol{x} \in E(\boldsymbol{z}(\boldsymbol{x}^{\text{obs}}))$ follows the truncated normal distribution with truncation intervals $E(\boldsymbol{z}(\boldsymbol{x}^{\text{obs}}))$, i.e.,

$$\left[\Delta \mid \Delta \in E(\boldsymbol{z}(\boldsymbol{x}^{\text{obs}}))\right] \sim \text{TN}(0, \boldsymbol{\eta}^\top \Sigma \boldsymbol{\eta}, E(\boldsymbol{z}(\boldsymbol{x}^{\text{obs}}))), \tag{5}$$

where $\text{TN}(\mu, \sigma^2, E)$ indicates the truncated normal distribution with mean μ, variance σ^2, and truncation intervals E. From (5),

$$\left[\Delta \,\middle|\, \begin{array}{c} \mathcal{A}(\boldsymbol{x}) = \mathcal{A}(\boldsymbol{x}^{\text{obs}}) \\ \boldsymbol{z}(\boldsymbol{x}) = \boldsymbol{z}(\boldsymbol{x}^{\text{obs}}) \\ \text{sgn}(\delta) = \text{sgn}(\delta^{\text{obs}}) \end{array}\right] \sim \text{TN}(0, \boldsymbol{\eta}^\top \Sigma \boldsymbol{\eta}, E(\boldsymbol{z}(\boldsymbol{x}^{\text{obs}}))).$$

This means that

$$\left[F_{0, \boldsymbol{\eta}^\top \Sigma \boldsymbol{\eta}}^{E(\boldsymbol{z}(\boldsymbol{x}^{\text{obs}}))}(\Delta) \,\middle|\, \begin{array}{c} \mathcal{A}(\boldsymbol{x}) = \mathcal{A}(\boldsymbol{x}^{\text{obs}}) \\ \boldsymbol{z}(\boldsymbol{x}) = \boldsymbol{z}(\boldsymbol{x}^{\text{obs}}) \\ \text{sgn}(\delta) = \text{sgn}(\delta^{\text{obs}}) \end{array}\right] \sim \text{Unif}[0, 1], \tag{6}$$

where F_{μ, σ^2}^{E} is the cumulative distribution function of the truncated normal distribution $\text{TN}(\mu, \sigma^2, E)$. This indicates property (4) and hence the validity of the proposed p-value computation method in (3).

C GC-based segmentation event

As stated in §3.2, the entire process of a maximum flow optimization problem can be decomposed into additions, subtractions, and comparisons of the weights

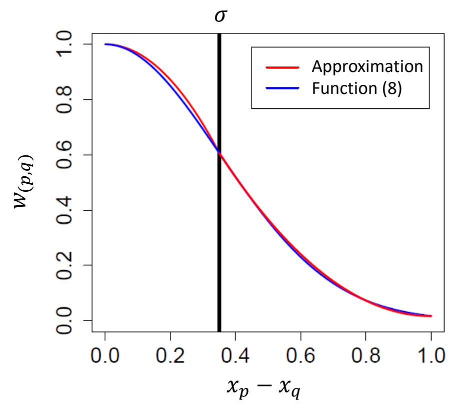


Figure 2: Example of a quadratic spline approximation of the commonly used weight function in (7).

assigned to the edges of the graph. Thus, if each weight is characterized by quadratic equations and inequalities on the image $\boldsymbol{x}$, the entire segmentation process can be represented by a set of quadratic inequalities in the form of (6) in the main text.

Recall that the graph contains $n+2$ nodes corresponding to n pixels and two terminal nodes S and T. The weight between two adjacent pixels is determined based on the similarity of their pixel intensities and the distance between them. Pixel similarity is usually defined based on the properties of the target image. To provide flexible choice of the similarity function, we employ a quadratic spline approximation, which allows one to specify the desired similarity function with arbitrary approximation accuracy. In the experiments in §4, we used a quadratic spline approximation of the commonly used weight function

$$w_{(p,q)} = \exp\left(-\frac{(x_p - x_q)^2}{2\sigma^2}\right)\frac{1}{\mathrm{dist}(p,q)}, (p,q) \in \mathcal{N}, \tag{7}$$

as shown in Figure 2.

In the rest of this section, we demonstrate that for a case with a quadratic spline approximation of (7), all the weights in the graph can be characterized by quadratic functions and inequalities on $\boldsymbol{x}$. When other similarity functions are used, if an appropriate quadratic spline approximation of the similarity function is employed, similar results can be obtained.

- $(p,q) \in \mathcal{N}$

$$w_{(p,q)} = \begin{cases} g_1(x_p - x_q)^2 + h_1, & \text{if } (x_p - x_q)^2 \leq \sigma^2, \\ g_2(x_q - x_p - 1)^2 + h_2, & \text{if } (x_p - x_q)^2 > \sigma^2, x_p \leq x_q, \\ g_2(x_p - x_q - 1)^2 + h_2, & \text{if } (x_p - x_q)^2 > \sigma^2, x_p > x_q, \end{cases} \tag{8}$$

where

$$g_1 = \frac{\exp(-\frac{1}{2}) - 1}{\sigma^2 \mathrm{dist}(p,q)}, \qquad h_1 = \frac{1}{\mathrm{dist}(p,q)},$$
$$g_2 = \frac{\exp(-\frac{1}{2}) - \exp(-\frac{1}{2\sigma^2})}{(\sigma-1)^2 \mathrm{dist}(p,q)}, h_2 = \exp\left(-\frac{1}{2\sigma^2}\right)\frac{1}{\mathrm{dist}(p,q)}.$$

The first inequality is written as $g_1(x_p - x_q)^2 + h_1 = \boldsymbol{x}^\top A_j \boldsymbol{x} + c_j$ with $A_j = g_1(\boldsymbol{e}_p - \boldsymbol{e}_q)(\boldsymbol{e}_p - \boldsymbol{e}_q), c_j = h_1$. The second quadratic function is written as $g_2(x_q - x_p - 1)^2 + h_2 = \boldsymbol{x}^\top A_j \boldsymbol{x} + \boldsymbol{b}_j^\top \boldsymbol{x} + c_j$ with $A_j = g_2(\boldsymbol{e}_p - \boldsymbol{e}_q)(\boldsymbol{e}_p - \boldsymbol{e}_q), \boldsymbol{b}_j = -2g_2(\boldsymbol{e}_q - \boldsymbol{e}_p), c_j = g_2 + h_2$. The third quadratic function is written as $g_2(x_p - x_q - 1)^2 + h_2 = \boldsymbol{x}^\top A_j \boldsymbol{x} + \boldsymbol{b}_j^\top \boldsymbol{x} + c_j$ with $A_j = g_2(\boldsymbol{e}_p - \boldsymbol{e}_q)(\boldsymbol{e}_p - \boldsymbol{e}_q), \boldsymbol{b}_j = -2g_2(\boldsymbol{e}_p - \boldsymbol{e}_q), c_j = g_2 + h_2$. The quadratic inequalities in the condition part are written as $(x_p - x_q)^2 \leq \sigma^2 \Leftrightarrow \boldsymbol{x}^\top A_j \boldsymbol{x} \leq 0$ and $(x_p - x_q)^2 > \sigma^2 \Leftrightarrow \boldsymbol{x}^\top A_j \boldsymbol{x} > 0$ with $A_j = (\boldsymbol{e}_p - \boldsymbol{e}_q)(\boldsymbol{e}_p - \boldsymbol{e}_q)^\top - \frac{1}{n-1}(I_n - n^{-1}\boldsymbol{e}_{\mathcal{P}}\boldsymbol{e}_{\mathcal{P}}^\top)$. The linear inequalities in the condition part are written as $x_p \leq x_q \Leftrightarrow \boldsymbol{b}_j^\top \boldsymbol{x} \leq 0$ and $x_p > x_q \Leftrightarrow \boldsymbol{b}_j^\top \boldsymbol{x} > 0$ with $\boldsymbol{b}_j = \boldsymbol{e}_p - \boldsymbol{e}_q$.

- $p = S, q \in \mathcal{P} \setminus (\mathcal{O}^{\mathrm{se}} \cup \mathcal{B}^{\mathrm{se}})$

$$w_{(S,q)} = \lambda \log(2\pi\sigma_{\mathrm{ob}}^2) + \frac{(x_q - m_{\mathrm{ob}}^{\mathrm{se}})^2}{2\sigma_{\mathrm{ob}}^2} \tag{9}$$

Noting that $m_{\mathrm{ob}}^{\mathrm{se}}$ is a linear function of $\boldsymbol{x}$ and assuming that σ_{ob}^2 is known or independently estimated as before, the weight in (9) is written as $\boldsymbol{x}_j^\top A_j \boldsymbol{x} + c_j$ with $A_j = \frac{\lambda}{2\sigma_{\mathrm{ob}}^2}(\boldsymbol{e}_p - \boldsymbol{e}_{\mathcal{O}^{\mathrm{se}}}/|\mathcal{O}^{\mathrm{se}}|)(\boldsymbol{e}_p - \boldsymbol{e}_{\mathcal{O}^{\mathrm{se}}}/|\mathcal{O}^{\mathrm{se}}|)^\top, c_j = \log(2\pi\sigma_{\mathrm{ob}}^2)$.

- $p = S, q \in \mathcal{O}^{\mathrm{se}}$

$$w_{(S,q)} = 1 + \max_{p \in \mathcal{P}} \sum_{r:(p,r) \in \mathcal{N}} w_{(p,r)}. \tag{10}$$

Let $k_p = \sum_{r:(p,r)\in\mathcal{N}} w_{(p,r)}$ for $p \in \mathcal{P}$. Since k_p is the sum of the weights characterized by quadratic functions and inequalities, as in (8), k_p is also characterized by quadratic functions and inequalities. Noting that $k_{\max} = \max_{p\in\mathcal{P}} k_p$ is characterized by a set of inequalities $k_{\max} \geq k$ for any $k \in \mathcal{P} \setminus k_{\max}$, the weight in (10), i.e., $k_{\max}$, is also characterized by quadratic functions and inequalities.

- $p = S, q \in \mathcal{B}^{\mathrm{se}}$

$$w_{(S,q)} = 0. \tag{11}$$

- $p \in \mathcal{P} \setminus (\mathcal{O}^{\mathrm{se}} \cup \mathcal{B}^{\mathrm{se}}), q = T$

$$w_{(p,T)} = \lambda \log(2\pi\sigma_{\mathrm{bg}}^2) + \frac{(x_q - m_{\mathrm{bg}}^{\mathrm{se}})^2}{2\sigma_{\mathrm{bg}}^2} \tag{12}$$

  As done for (9), the weight in (12) can be written as $\boldsymbol{x}_j^\top A_j \boldsymbol{x} + c_j$ with $A_j = \frac{\lambda}{2\sigma_{\mathrm{bg}}^2}(\boldsymbol{e}_p - \boldsymbol{e}_{\mathcal{B}^{\mathrm{se}}}/|\mathcal{B}^{\mathrm{se}}|)(\boldsymbol{e}_p - \boldsymbol{e}_{\mathcal{B}^{\mathrm{se}}}/|\mathcal{B}^{\mathrm{se}}|)^\top, c_j = \log(2\pi\sigma_{\mathrm{bg}}^2)$.

- $p \in \mathcal{O}^{\mathrm{se}}, q = T$

$$w_{(p,T)} = 1 + \max_{q \in \mathcal{P}} \sum_{r:(r,q) \in \mathcal{N}} w_{(r,q)}. \tag{13}$$

  As done for (10), the weight in (13) can be characterized by quadratic functions and inequalities.

- $p \in \mathcal{B}^{\mathrm{se}}, q = T$

$$w_{(p,T)} = 0. \tag{14}$$