

ขั้นตอนวิธีเชิงตัวเลขชนิดใหม่สำหรับการต่อเติมภาพที่ใช้การแปรผันรวมกับการประยุกต์
สำหรับซ่อมแซมภาพจิตรกรรมไทยโบราณและการลบบทบรรยายจากอนิเมะ

A new numerical algorithm for TV-based image inpainting with its applications
for restoring ancient Thai painting images and removing subtitles from animes

ภัคพล พงษ์ทวี

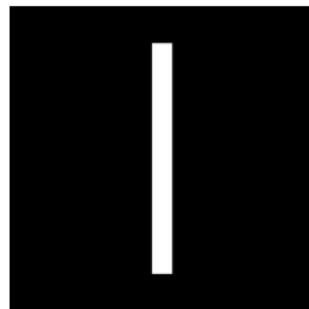
ภาควิชาคณิตศาสตร์
มหาวิทยาลัยศิลปากร

งานประชุมวิชาการสำหรับนักศึกษาและดับเบิลยูญาตรี
สาขาวิชาคณิตศาสตร์ประยุกต์ ครั้งที่ 8
27 เมษายน 2562

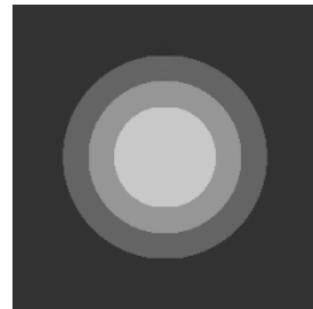
การต่อเติมภาพ (Image Inpainting)



(a) ภาพที่ต้องการซ่อมแซม



(b) โฉมหนต่อเติม



(c) ภาพที่ได้รับการซ่อมแซม

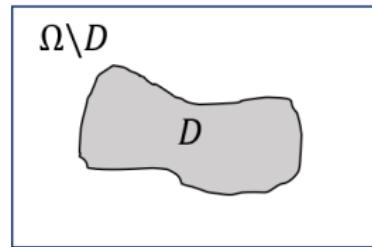
รูปที่ 1: ตัวอย่างการซ่อมแซมภาพ

ปัญหาการต่อเติมภาพ

- (1) การต่อเติมภาพแนวเทา
- (2) การต่อเติมภาพสี

ภาพฉาดเทา

- โดเมนภาพ (image domain) $\Omega \subset \mathbb{R}^2$
- โดเมนต่อเติม (inpainting domain) $D \subset \mathbb{R}^2$
- พิกัดทางกายภาพ (physical position) $\mathbf{x} = (x, y) \in \Omega$
- ระดับความเข้มของภาพ (image intensity) $V \subset [0, \infty)$
- ภาพฉาดเทา (grayscale image) $u : \Omega \rightarrow V$, $z : \Omega \rightarrow V$
- โดยไม่เสียหลักการสำคัญ $\Omega = [1, n]^2$ และ $V = [0, 1]$ เมื่อ $n > 0$ เป็นจำนวนเต็มบวก



รูปที่ 2: D แทนโดเมนต่อเติม

ตัวแบบการต่อเติมภาพเขตสีเทาที่ใช้การแปรผันรวม

$$\min_u \{ \mathcal{J}(u) = \frac{1}{2} \int_{\Omega} \lambda(u - z)^2 d\Omega + \int_{\Omega} |\nabla u| d\Omega \}$$

$$\lambda = \lambda(x) = \begin{cases} \lambda_0, & x \in \Omega \setminus D \\ 0, & x \in D \end{cases}$$

T.F. Chan and J. Shen , "Mathematical models of local non-texture inpaintings", SIAM Journal on Applied Mathematics, vol. 62, no. 3, pp. 1019–1043, 2001.

ตัวแบบการต่อเติมภาพเอดสีเทาที่ใช้การแปรผันรวม (ต่อ)

$$\min_u \{ \mathcal{J}(u) = \frac{1}{2} \int_{\Omega} \lambda(u - z)^2 d\Omega + \int_{\Omega} |\nabla u| d\Omega \}$$



$$\begin{cases} -\nabla \cdot \left(\frac{\nabla u}{|\nabla u|} \right) + \lambda(u - z) = 0, & \mathbf{x} \in (1, n)^2 \\ \frac{\partial u}{\partial n} = 0, & x \in \partial\Omega \end{cases}$$

วิธีการเชิงตัวเลขสำหรับการกำหนดสัญญาณรบกวน

- (1) การเดินเวลาแบบชัดเจ้ง
- (2) การทำข้าแบบจุดตึง
- (3) วิธีการสปริทเบรกเมน

การเดินเวลาแบบชัดแจ้ง (explicit time marching)

$$u(\mathbf{x}, t_{k+1}) = u(\mathbf{x}, t_k) + \tau \left(\nabla \cdot \left(\frac{\nabla u(\mathbf{x}, t_k)}{|\nabla u(\mathbf{x}, t_k)|} \right) + \lambda(\mathbf{x})(u(\mathbf{x}, t_k) - z(\mathbf{x})) \right)$$

$$u(\mathbf{x}, t_0) = z \quad t_k = t_0 + k\tau \ (\tau > 0) \quad t_0 = 0$$

$$u(\mathbf{x}, t_0), u(\mathbf{x}, t_1), u(\mathbf{x}, t_2), u(\mathbf{x}, t_3), \dots, u(\mathbf{x}, t^*)$$

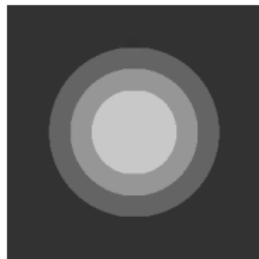
L. I. Rudin, S. Osher, E. Fatemi, "Nonlinear total variation based noise removal algorithms", Physica D: Nonlinear Phenomena, vol 60, issues 1–4, pp. 259–268, 1992.

การทำซ้ำแบบจุดตรึง (fixed-point iteration)

$$-\nabla \cdot \left(\frac{\nabla u^{[\nu+1]}}{|\nabla u|^{[\nu]}} \right) + \lambda(u^{[\nu+1]} - z) = 0, \quad u^{[0]} = z$$

$u^{[0]}, u^{[1]}, u^{[2]}, u^{[3]}, \dots, u^*$

ปัญหาเชิงตัวเลข



รูปที่ 3: ตัวอย่างภาพที่เกิดปัญหาเชิงตัวเลข

$$\frac{1}{|\nabla u|} = \frac{1}{\sqrt{u_x^2 + u_y^2}} \rightarrow \infty$$

$$|\nabla u| \approx |\nabla u|_\beta = \sqrt{u_x^2 + u_y^2 + \beta}, \quad 0 < \beta \ll 1$$

วิธีการสปริทเบรกแม่น (Split Bregman method)

$$\min_{u,w} \{ \mathcal{J}(u, w) = \frac{1}{2} \int_{\Omega} \lambda(u - z)^2 d\Omega + \int_{\Omega} |\nabla w| d\Omega + \frac{\theta}{2} \int_{\Omega} (w - \nabla u + b) d\Omega \}$$

T. Goldstein and S. Osher, "The Split Bregman Method for L1-Regularized Problems", SIAM Journal on Imaging Sciences. vol. 2, issue 2, pp. 323-343, 2009.

วิธีการสปริทเบรคแม่น (ต่อ)

$$\min_{u,w} \{ \mathcal{J}(u, w) = \frac{1}{2} \int_{\Omega} \lambda(u - z)^2 d\Omega + \int_{\Omega} |\nabla w| d\Omega + \frac{\theta}{2} \int_{\Omega} (w - \nabla u + b) d\Omega \}$$

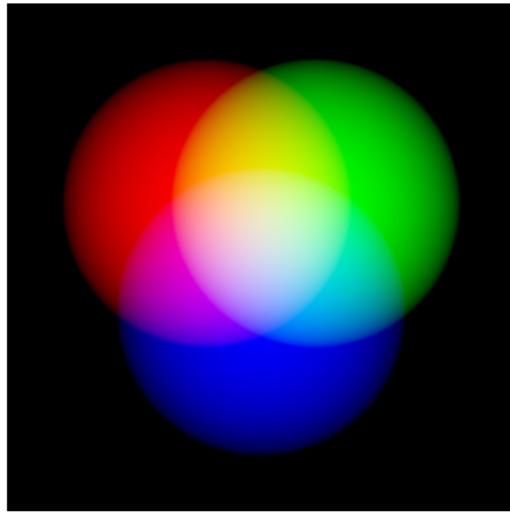


$$u^{\text{New}} = \arg \min_u \{ \mathcal{J}_1(u) = \frac{1}{2} \int_{\Omega} \lambda(u - z)^2 d\Omega + \frac{\theta}{2} \int_{\Omega} (w^{\text{old}} - \nabla u + b^{\text{old}}) d\Omega \}$$

$$w^{\text{New}} = \arg \min_w \{ \mathcal{J}_2(w) = \int_{\Omega} |\nabla w| d\Omega + \frac{\theta}{2} \int_{\Omega} (w - \nabla u^{\text{New}} + b^{\text{old}}) d\Omega \}$$

$$b^{\text{New}} = b^{\text{old}} + \nabla u^{\text{New}} - w^{\text{New}}$$

ภาพสี



รูปที่ 4: ภาพสีเกิดจากการผสมกันระหว่างสี แดง เขียว และน้ำเงิน¹

¹ ภาพจาก https://commons.wikimedia.org/wiki/File:Additive_RGB_Circles+48bpp.png สืบคันมือวันที่ 25 กันยายน 2561

ກາພສີ (ຕ່ອ)

$u, z : \Omega \rightarrow V$



$$u = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}, z = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} : \Omega \rightarrow V^3$$

ตัวแบบการต่อเติมภาพสีที่ใช้การแปรผันรวม

$$\min_u \{ \mathcal{J}(u) = \frac{1}{2} \int_{\Omega} \lambda(u - z)^2 d\Omega + \int_{\Omega} |\nabla u| d\Omega \}$$



$$\min_u \{ \mathcal{J}(u) = \sum_{l=1}^3 (\frac{1}{2} \int_{\Omega} \lambda(u_l - z_l)^2 d\Omega + \int_{\Omega} |\nabla u_l| d\Omega) \}$$

วิธีการสปริทเบรคแม่นสำหรับภาพสี

$$\min_{u,w} \{ \mathcal{J}(u, w) = \frac{1}{2} \int_{\Omega} \lambda(u - z)^2 d\Omega + \int_{\Omega} |\nabla w| d\Omega + \frac{\theta}{2} \int_{\Omega} (w - \nabla u + b) d\Omega \}$$



$$\begin{aligned} \min_{u, w_1, w_2, w_3} \{ \mathcal{J}(u, w_1, w_2, w_3) &= \sum_{l=1}^3 \left(\frac{1}{2} \int_{\Omega} \lambda(u_l - z_l)^2 d\Omega + \int_{\Omega} |\nabla w_l| d\Omega \right. \\ &\quad \left. + \frac{\theta}{2} \int_{\Omega} (w_l - \nabla u_l + b_l) d\Omega \right) \end{aligned}$$

Peak Signal Noise Ratio (PSNR)

$$\text{PSNR} = 10 \cdot \log_{10}\left(\frac{1}{\sqrt{\text{MSE}}}\right)$$

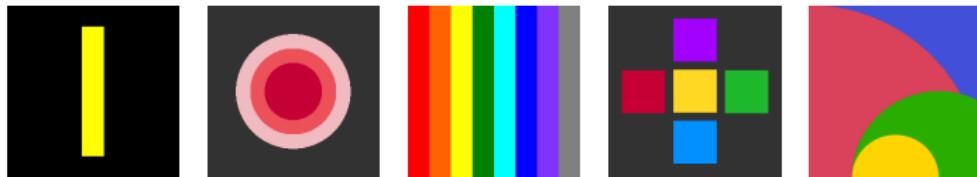
- MSE คือค่าคลาดเคลื่อนกำลังสองเฉลี่ยของภาพ โดยที่ $\text{MSE} = \frac{1}{nx \times ny} \sum (u - \tilde{u})^2$
- u แทนภาพต้นฉบับ
- \tilde{u} แทนภาพที่ได้จากการซ่อมแซมโดยวิธีเชิงตัวเลข
- PSNR มีหน่วยเป็น dB

Structural Similarity (SSIM)

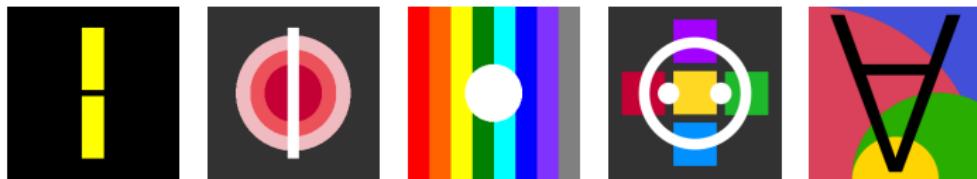
$$\text{SSIM}(u, \tilde{u}) = \frac{(2\mu_u\mu_{\tilde{u}} + 0.0001)(2\sigma_{u\tilde{u}} + 0.0009)}{(\mu_u^2 + \mu_{\tilde{u}}^2 + 0.0001)(\sigma_u^2 + \sigma_{\tilde{u}}^2 + 0.0009)}$$

- u แทนภาพต้นฉบับ
- \tilde{u} แทนภาพต้นฉบับ และภาพที่ได้จากการซ่อมแซมโดยวิธีเชิงตัวเลข
- μ_u คือค่าเฉลี่ยของ u
- $\mu_{\tilde{u}}$ คือค่าเฉลี่ยของ \tilde{u}
- σ_u คือความแปรปรวนของ u
- $\sigma_{\tilde{u}}$ คือความแปรปรวนของ \tilde{u}

ภาพสังเคราะห์



รูปที่ 5: ภาพต้นฉบับ

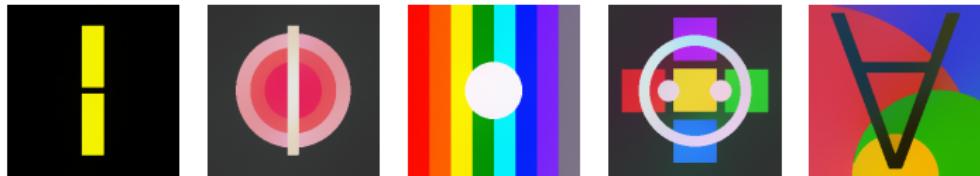


รูปที่ 6: ภาพที่จะทำการซ่อมแซม

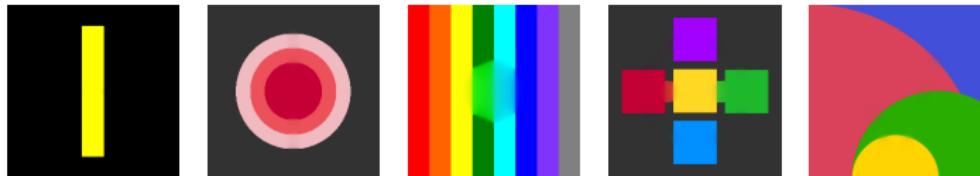
รอบการทำซ้ำ $\leq 10,000$ รอบ

$$\frac{\|u_{new} - u_{old}\|}{\|u_{new}\|} \geq 10^{-4}$$

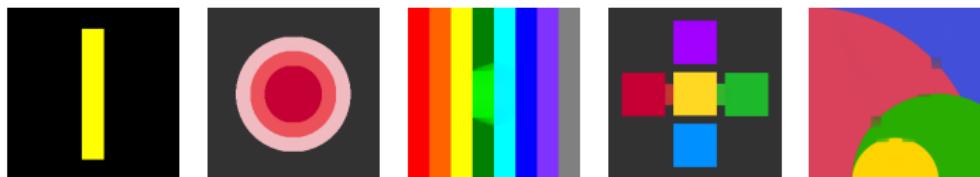
ผลการต่อเติมภาพ



รูปที่ 7: ผลการซ่อมแซมจากการเดินเวลา



รูปที่ 8: ผลการซ่อมแซมจากการทำข้าวแบบจุดตรง



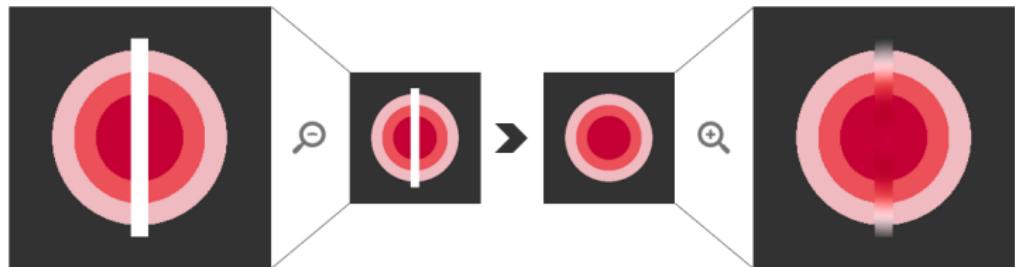
รูปที่ 9: ผลการซ่อมแซมจากการสปริทเบรกแม่น

ประสิทธิภาพของวิธีการเชิงตัวเลขทั้ง 3 วิธี

วิธีการ	เวลาประมาณ (วินาที)	PSNR (dB)	SSIM
การเดินเวลา	120.68	16.72	0.9960
การทำซ้ำจุดตรึง	74.81	38.67	0.9999
การสปริทเบรกเมน	14.06	39.42	0.9999

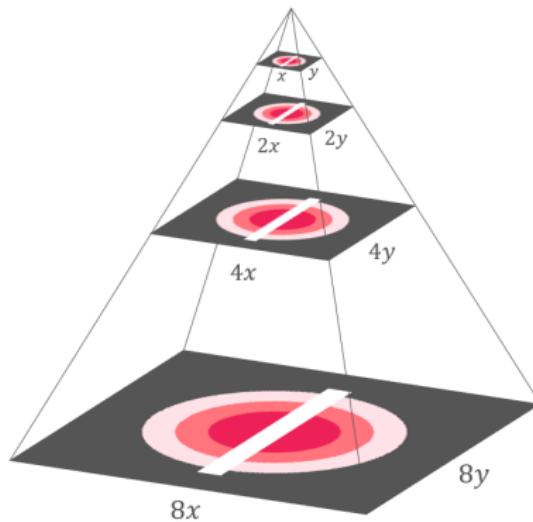
ตารางที่ 1: แสดงการซ่อมแซมเฉลี่ยของวิธีการเชิงตัวเลข
โดยที่ $\lambda = 250, \beta = 10^{-5}, \tau = 10^{-5}, \theta = 5$

คำตอบเริ่มต้น (initial solution)



รูปที่ 10: สร้างคำตอบเริ่มต้นโดยใช้พีระมิดรูปภาพ

พีระมิดรูปภาพ



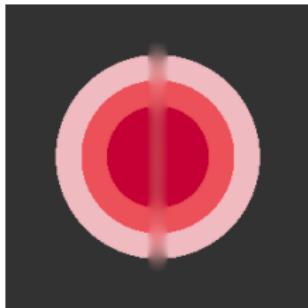
รูปที่ 11: วิธีการพีระมิดรูปภาพ

ผลการซ่อมแซมเมื่อใช้พิระมิดรูปภาพ

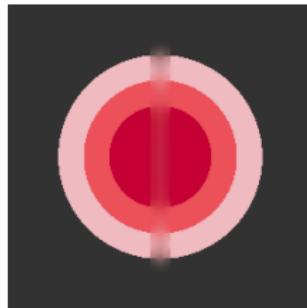
รูปแบบการทำซ้ำ	เวลาประมาณ (วินาที)	PSNR (dB)	SSIM
ไม่ใช้พิระมิดรูปภาพ	17.38	39.42	0.9999
10/1/1/10000	13.52	39.38	0.9999
10/3/3/10000	11.86	39.54	0.9999
10/10/10/10000	9.26	40.17	0.9999
100/1/1/10000	10.28	39.04	0.9999
100/3/3/10000	10.28	39.80	0.9999
100/10/10/10000	9.27	40.12	0.9999

ตารางที่ 2: ผลการซ่อมแซมภาพโดยวิธีการเขิงตัวเลขที่นำเสนอในรูปของค่าเฉลี่ย

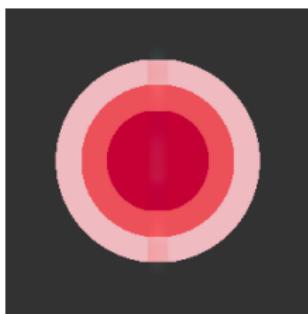
การทำซ้ำความละเอียดคอมพิวเตอร์สูง



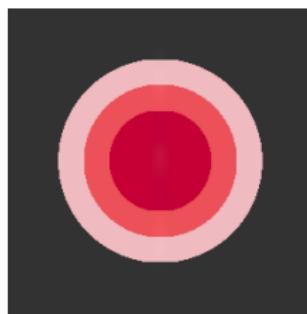
(a) 5 ครั้ง



(b) 10 ครั้ง



(c) 50 ครั้ง



(d) 100 ครั้ง

รูปที่ 12: พิริมิดที่ลำดับการทำซ้ำเป็น 10/10/10 และที่ระดับความคมชัดละเอียดสุดใช้จำนวนการทำซ้ำที่ต่างกัน

ผลการซ่อมแซมภาพในระดับความคมชัดละเอียดสุด 10 ครั้ง

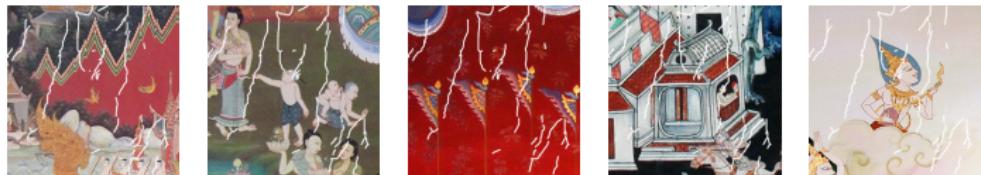
รูปแบบการทำข้าว	เวลาประมาณ (วินาที)	PSNR (dB)	SSIM
ไม่ใช้พรมมิตรรูปภาพ	0.37	17.26	0.9963
10/1/1/10	0.40	28.54	0.9993
10/3/3/10	0.33	29.83	0.9994
10/10/10/10	0.38	32.56	0.9995
100/1/1/10	0.34	31.50	0.9999
100/3/3/10	0.36	31.99	0.9999
100/10/10/10	0.38	33.39	0.9998

ตารางที่ 3: ผลการซ่อมแซมภาพเฉลี่ยเมื่อใช้การทำข้าวในระดับความคมชัดละเอียดสุด 10 ครั้ง

การซ่อมแซมภาพจิตรกรรมไทยโบราณ



รูปที่ 13: ภาพต้นฉบับสำหรับใช้ในการทดสอบ



รูปที่ 14: ภาพที่ทำให้เสียหาย

ผลการซ่อมแซมภาพศิลปะไทย



รูปที่ 15: ผลการซ่อมแซมโดยวิธีการสบริทเบรกแม่น



รูปที่ 16: ผลการซ่อมแซมโดยวิธีการที่พัฒนาขึ้น

เปรียบเทียบระหว่าง 2 วิธี



(a) ภาพต้นฉบับ



(b) ภาพที่เสียหาย



(c) วิธีสปริทเบรกแม่น



(d) วิธีที่พัฒนาขึ้น

รูปที่ 17: เปรียบเทียบจากภาพที่ถูกขยายขึ้น 4 เท่า

ผลการซ่อมแซมภาพศิลปะไทย

วิธีการ	เวลาประมาณ (วินาที)	PSNR (dB)	SSIM
สเปรย์เบรกเม็น	2.72	34.89	1.0000
วิธีการที่พัฒนาขึ้น	0.39	35.30	1.0000

ตารางที่ 4: แสดงผลการซ่อมแซมภาพศิลปะไทยในรูปค่าเฉลี่ย

การลับบทบรรยายบนอนิเมะ



รูปที่ 18: Festival Asia Special Video - feat. Inori Aizawa

Lorem Ipsum

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat.

But I must explain to you how all this mistaken idea of denouncing of a pleasure and praising pain was born and I will not give you a complete account of the system, and expound the actual teachings of the great explorer of the truth, the master-builder of human happiness. No one rejects, dislikes, or avoids pleasure itself, because it is pleasure

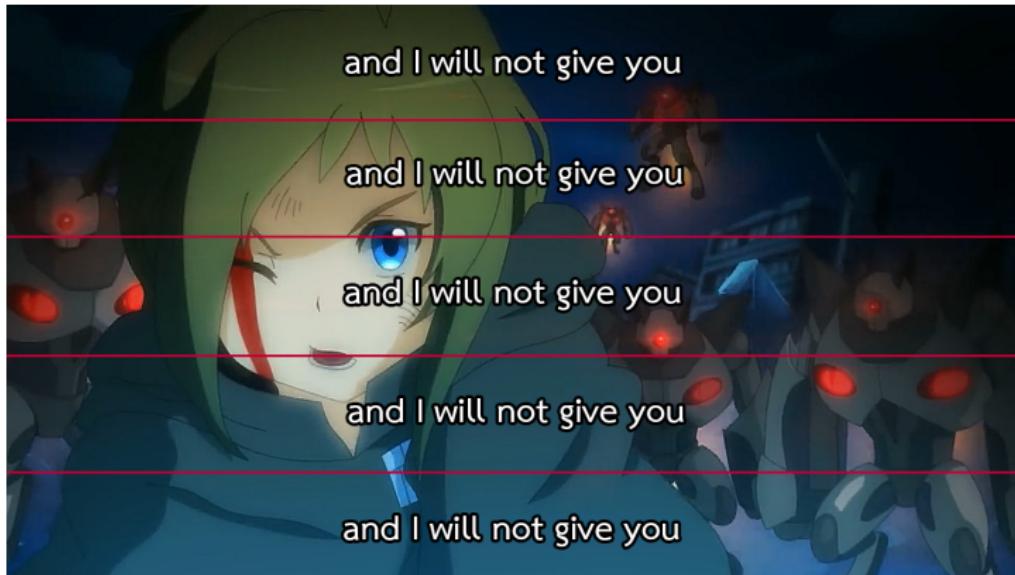
しかし私は、喜びを非難して苦痛を賞賛するという誤ったこの考えがすべてどのようにして誕生したかをあなたに説明しなければならないから、私はあなたにその体系を完璧に説明し、真実を求める偉大な探究家、人間の喜びを築く建築家の実践的な教えを詳しく説明しよう

*หากแต่ข้าพเจ้าต้องขออภัยต่อท่านในเหตุที่มิโนคิดอันความชิงชักอันหลงผิดทั้งหลาย ในการไฟห้าแฉะสุดีบัดແພດ
 ว่าบ่มเพาะขึ้นได้อย่างไรแล้ว ข้าพเจ้าจักให้สาระแห่งระบบอันครบถ้วนแด่ท่าน และประสิทธิ์คำสอนที่แท้แห่งการแสวงหาอัน
 ยิ่งใหญ่ในความจริง ซึ่งเป็นบูรพคณาจารย์ผู้สร้างร่างความมาสุขแห่งมวลมนุษย์ ทุกจะมีผู้ได้มิรับ มิชอบ หรือเลี้ยงหลีก
 ความสุข ด้วยเพราเวณเป็นสุขก็หายไป*

รูปที่ 19: Lorem Ipsum ทั้ง 4 ภาษา²

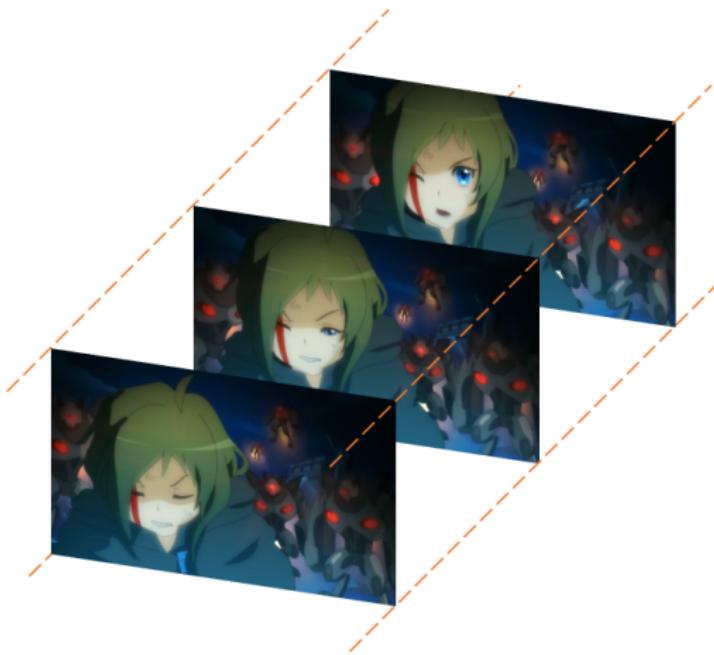
² ส่วนหนึ่งของ Lorem ipsum ในภาษาต่างๆ จาก https://en.wikipedia.org/wiki/Lorem_ipsum, https://ja.wikipedia.org/wiki/Lorem_ipsum และ <https://th.wikipedia.org/wiki/ລອມເມື່ອສັນເກີດ> สืบค้นเมื่อ 25 พฤษภาคม 2561

การแบ่งไฟล์วิดีโอ



รูปที่ 20: แบ่งวิดีโอออกเป็น 5 ส่วน

วิดีโอกับภาพ



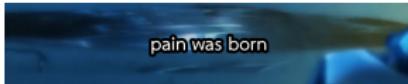
รูปที่ 21: วิดีโอกีฬาดับเบิลของภาพ

การห้าบทบรรยาย

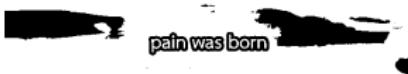
This is anime subtitle

รูปที่ 22: บทบรรยายอนิเมะมักมีขอบเป็นสีดำ

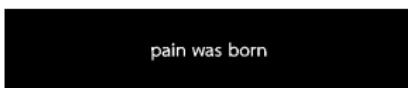
การหาบทบรรยาย (ต่อ)



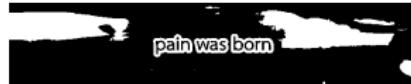
(a) ภาพเฟรมอนิเมะบริเวณที่มีบทบรรยาย



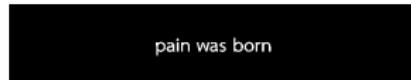
(c) สลับสีขาวเป็นสีดำ สีดำเป็นสีขาว



(e) ลบสีที่เล็กเกินไปหรือใหญ่เกินไป



(b) ให้สีดำในภาพเป็นสีขาว สีอื่นเป็นสีดำ



(d) ภาพหลังทำการเปลี่ยนพื้นที่สีขาว



(f) ขยายโดเมนต่อเติมให้ครอบคลุม

รูปที่ 23: ขั้นตอนวิธีการหาบทบรรยาย

ผลลัพธ์การ hababrery

ภาษา	จำนวนพิกเซลในโดเมน	จำนวนพิกเซลที่ตรวจพบ	จำนวนพิกเซลที่ผิดพลาด	ร้อยละการผิดพลาด
ไทย	23,222,220	24,083,125	2,141,201	9.22
อังกฤษ	27,278,745	28,598,424	3,714,321	13.62
ญี่ปุ่น	28,544,173	30,103,466	3,740,971	13.11

ตารางที่ 5: ความคลาดเคลื่อนเฉลี่ยของการหาโดเมนต่อitem ในบทบรรยายภาษาต่างๆ

ลำดับภาพกับคำตอบเริ่มต้น

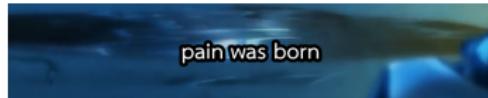


รูปที่ 24: ลำดับภาพในวิดีโอบริเวณคำบรรยาย

พิจารณาความคล้ายด้วย SSIM



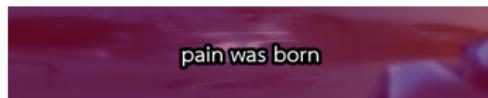
(a) เพรมก่อนหน้า



(b) เพรมปัจจุบัน



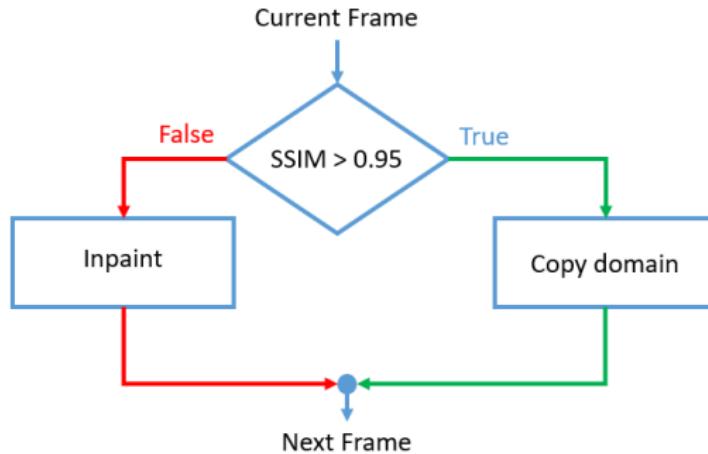
(c) บริเวณคำนวณ SSIM ของเพรมก่อนหน้า



(d) บริเวณคำนวณ SSIM ของเพรมปัจจุบัน

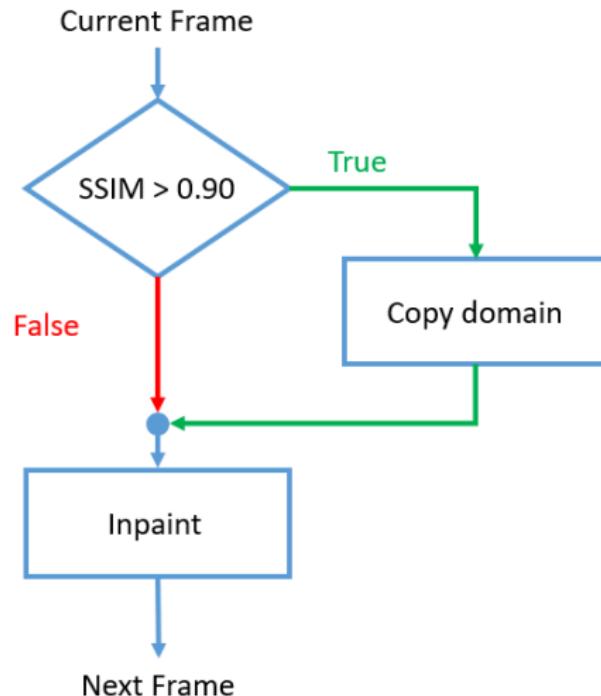
รูปที่ 25: บริเวณที่คำนวณ SSIM สำหรับการข้ามเพรมและยีมเพรม

การข้ามเฟรม



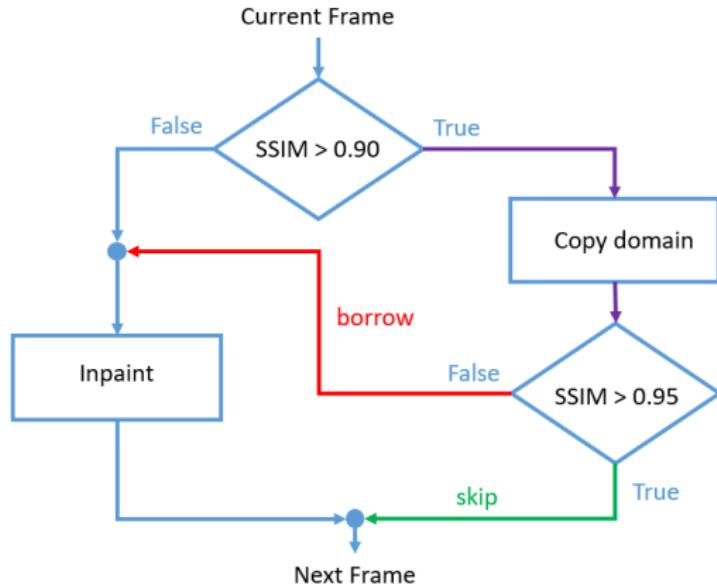
รูปที่ 26: ขั้นตอนวิธีการข้ามเฟรม

การรีบูตเฟรม



รูปที่ 27: ขั้นตอนวิธีการรีบูตเฟรม

การข้ามเฟรมและยึดเฟรม



รูปที่ 28: ขั้นตอนวิธีการข้ามและยึดเฟรม

ผลของวิธีข้ามเฟรมและยึดเฟรม

วิธีการ	เวลาประมาณ (วินาที)	PSNR (dB)	SSIM
สปริทเบรกแม่นและพีระมิดรูปภาพ	141.29	31.39	0.9510
ข้ามเฟรม	89.29	29.07	0.9408
ยึดเฟรม	132.78	32.20	0.9655
ข้ามเฟรมและยึดเฟรม	75.76	29.33	0.9454

ตารางที่ 6: การลับคำบรรยายเฉลี่ยด้วยวิธีต่างๆ

ผลการลบคำบรรยาย



(a) ก่อนลบคำบรรยาย



(b) หลังลบคำบรรยาย

ผลการลบคำบรรยาย (ต่อ)

วิธีการ	เวลาประมวล (วินาที)	PSNR (dB)	SSIM
สปริทเบรกแม่น	5073.08	32.88	0.9654
วิธีการที่พัฒนาขึ้น	75.76	29.33	0.9454

ตารางที่ 7: ผลการลบบทบรรยายออกจากอนิเมะเฉลี่ย
โดยวิธีการสปริทเบรกแม่นและวิธีการที่พัฒนาขึ้น

ลองซ้อมแซมภาษาพศิลปะ

[]

ภาษาพศิลป์: ล้วงคำไว้ทุกรอย (๑)

เมื่อเลือกภาษาพศิลป์คือเป็นลักษณะภาษาที่ใช้บ้านบ้าน ทางเดิมก็เป็นไทยต้นมา จนปัจจุบันจะเรียกว่า “ไทยเดิม” หรือ “ภาษาไทยเดิม”

พารามิเตอร์สำหรับการต่อเนื่อง

สำหรับภาษาพศิลป์จะต้องใช้เวลาในการประมวลผลประมาณ 20 วินาที ที่สำคัญคือจะขอร้องภาษาต่อเนื่องมาให้ที่สุดไปไม่ถึงภาษาเดิมและกับบุคคลฯ ฝึกภาษาเดิมจะต้องมีความต้องการที่จะรับฟังภาษาเดิมและรับฟังภาษาเดิมในรูปแบบเดิมๆ ของภาษาเดิม เช่น คำที่มีเสียงสูง เช่น “น้ำ” คำที่มีเสียงต่ำ เช่น “น้ำ” คำที่มีเสียงสูง เช่น “น้ำ” คำที่มีเสียงต่ำ เช่น “น้ำ”

θ: 5

λ: 250

ε: 0.0001

ภาษาที่ร่างขึ้นและอ่าน: 10ภาษาที่ร่างขึ้นและหัวใจของ: 3ภาษาที่ร่างขึ้นและหมาย: 10ความลึกของพื้นที่: 4

⑤

ภาษาพศิลป์ที่ต้องการซ่อนแซม



รูปที่ 30: ลองใช้งานได้ที่ <https://bit.ly/demothai>

ขอขอบคุณ