

Deep learning によるシングルショット OCT 画像のノイズ除去

坂下祐輔^{1, 2)}, 熊谷佳紀¹⁾, 柴涼介¹⁾, 竹野直樹¹⁾, 山下隆義²⁾, 藤吉弘亘²⁾

¹⁾ 株式会社ニデック アイケア事業部 開発本部

²⁾ 中部大学

Image denoising from a single-shot of OCT using deep learning

Yusuke Sakashita^{1, 2)}, Yoshiki Kumagai¹⁾, Ryosuke Shiba¹⁾, Naoki Takeno¹⁾,
Takayoshi Yamashita²⁾, Hironobu Fujiyoshi²⁾

¹⁾ Development Div., Eye Care Div., NIDEK CO., LTD.

²⁾ Chubu University

光干渉断層計 (OCT) の発展に伴い, 眼底 OCT 画像の撮影範囲や画質が向上している. しかし, 眼底 OCT 画像はスペckルノイズ等のノイズが含まれるため, 加算平均処理によって画質を向上させている. 加算平均処理は, 同一部位を複数回スキャンし, 得られた複数の画像を加算することでノイズを低減する手法であるが, 加算枚数を増やすとそれに比例して撮影時間が増加し, 被検者への負担がかかる. 本稿では, シングルショットの OCT 画像から 120 枚加算平均相当の画像を取得することを目的とし, Deep learning によるシングルショット OCT 画像のノイズ除去を提案する. また, 客観的評価および主観的評価により, 提案手法の有効性を示す.

キーワード: 光干渉断層計, 深層学習, 敵対的生成ネットワーク, デノイジング

As optical coherence tomography (OCT) technology has developed, the photographing range and the image quality of fundus OCT images have improved. However, fundus OCT images contain noise such as speckle noise. Application of the averaging method will improve their image qualities. In the averaging technique, noise is reduced by scanning the same area multiple times and then averaging images to obtain a clear image. The more frames the better the image, but this requires more time for photographing, which causes a burden on patients. We proposed noise reduction from a single scan of OCT using Deep Learning, aiming to obtain images conforming to 120 images used by the averaging method from a single-shot of OCT. The effectiveness of the proposed method is shown through both objective and subjective evaluations.

Keywords: Optical coherence tomography, Deep learning, Generative adversarial network, Denoising

1. 緒言

光干渉断層計 (Optical Coherence Tomography, OCT) は網膜の断層面を非侵襲的かつ高速で撮影することが可能であり, 眼科における検査や検診にて広く使われている¹⁾.

一方で, OCT から得られる画像は, ノイズの発生や, 感度の減衰により不鮮明な画像になってしまうため, その対策が必要とされている. 一般的な対策として行われている処理が, 加算平均処理である²⁾ (図 1).

加算平均処理は, 同一部位を複数回スキャンし, 取

得した複数枚の画像を加算して平均する処理である. 一般に加算枚数 N 枚のとき SNR (signal-to-noise ratio) は $\sqrt{N}$ 倍になるため, ノイズが除去され, 網膜の各層の境界が鮮明な画像になる. 加算平均処理は, 加算枚数を増やせば増やすほど画像は鮮明になっていくが, それに比例して測定時間も増加し (A スキャン速度 85k, スキャンポイント 1024 の場合, 一般的にシングルショットで数十ミリ秒, 120 枚加算で理論値 6 秒以上), 被験者への負担が大きくなる. また, 測定中はトラッキング技術による位置補正が行われているものの, 安定した固視が必要であり, 疾患

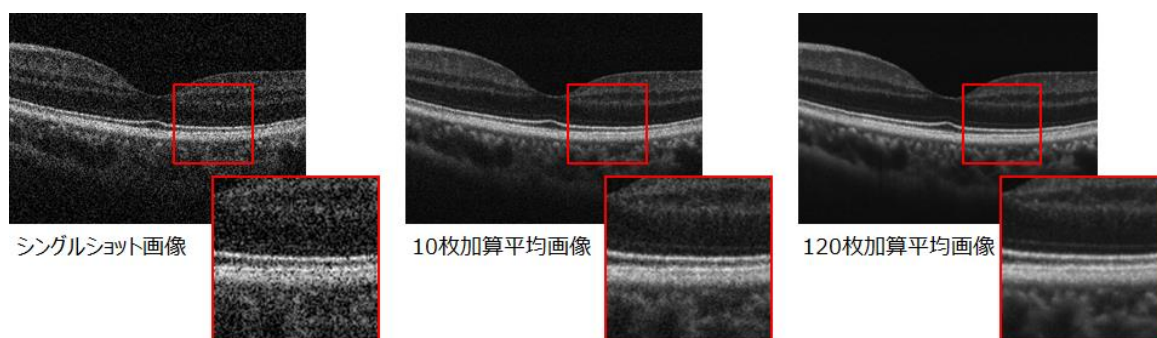


図 1 加算平均処理の効果

同じ被検者のシングルショット画像（加算平均なし）、10 枚による加算平均画像、120 枚による加算平均画像である。加算枚数を増やすほど、ノイズが軽減され鮮明な画像を取得することが可能となる。

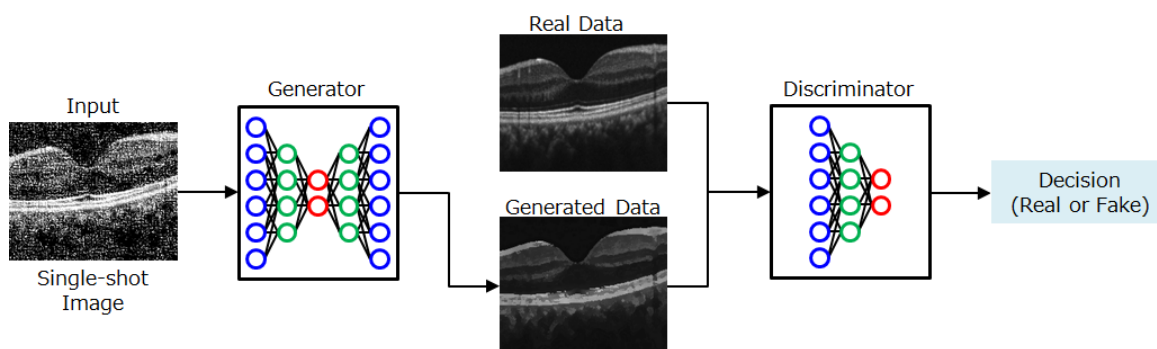


図 2 本手法の概要

Generator はシングルショット画像を入力とし、120 枚加算平均画像と見分けがつかない画像を出力するように学習し、Discriminator は、本物の 120 枚加算平均画像と Generator が出力した画像を正しく見分けるように学習していく。Generator と Discriminator が競い合うように交互に学習することで、Generator の生成能力と Discriminator の識別能力が向上していく。

を伴う被験者や高齢者での固視状態では、加算平均処理がうまく働かないことも考えられる。そのため、複数回スキャンを必要としない画像処理によるノイズ除去が求められる。これまで、1 枚の OCT 画像のノイズ除去としては、ガウシアンフィルタ等の平滑化手法が用いられてきた。しかし、これらの手法はノイズを抑制する効果はあるものの、見た目が不自然になる傾向にある。

このような問題を解決するため、近年では Convolutional Neural Network (CNN) を用いた画像のノイズ除去が成果を上げており、OCT 画像においても有効性が報告されている³⁾。

その一方で、CNN の学習において敵対的に学習を行う Generative adversarial network (GAN) の適用が、より鮮明な画像を生成するために有効な手法であることも報告されている⁴⁾。また、GAN は Computed Tomography (CT) 画像や Magnetic Resonance

(MR) 画像のノイズ除去、超解像処理で用いられ、医用画像に対しても有効性が報告されている^{5, 6)}。

そこで、本研究では GAN を用いてシングルショット OCT 画像からノイズを除去することを目的とし、その有効性を確認する。

2. 方法

2.1 概要

GAN は 2 つのネットワークから構成される。Generator と呼ばれる生成モデルと、Discriminator と呼ばれる識別モデルがあり、どちらも多層ニューラルネットワークの構造に基づいている。本手法の概要を図 2 に示す。

Generator は、シングルショット画像からノイズを除去した画像を出力することを目的としたネットワークであり、シングルショット画像と 120 枚加算平均画像のペアからなるデー

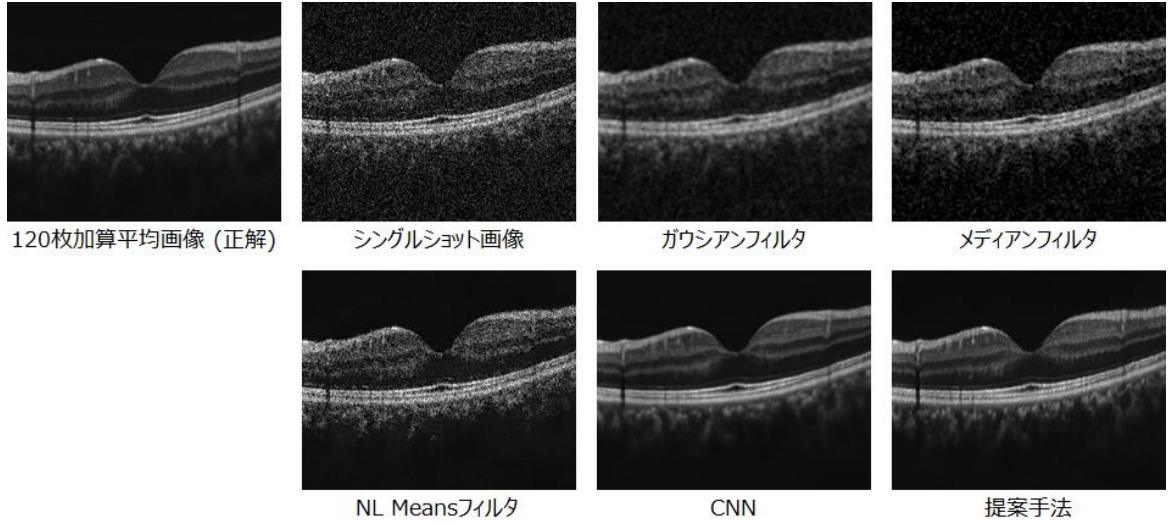


図4 シングルショット画像を入力としたノイズ除去結果
正解画像である120枚加算平均画像に対し、シングルショット画像、ガウシアンフィルタ、メディアンフィルタ、NL Means フィルタ、CNN、提案手法によるノイズ除去結果。

$$L_{\text{adversarial}} =$$

$$\min_G \max_D ([\log D(x)] + [\log(1 - D(G(i)))]]) \quad \dots\dots(1)$$

ここで、 D は Discriminator、 G は Generatorを表し、 i は入力画像、 x は本物の画像、 $G(i)$ は G が出力した画像である。Generator は式(1)が最小になるようにネットワークを更新し、Discriminator は式(1)が最大になるように交互にネットワークを更新していく。

一方、提案手法の Generator の学習に用いる目的関数は、式(1)で記述される Adversarial loss に加え、SSIM loss を加えた式(2)を用いる。

$$Loss_{\text{Generator}} = L_{\text{adversarial}} + \lambda \cdot L_{\text{SSIM}} \quad \dots\dots(2)$$

ここで、 λ は SSIM loss に対する重みを表し、本手法では $\lambda = 2.0$ とした。SSIM (Structural Similarity) は両画像間の構造的な類似を求める指標であり⁸⁾、CNN の目的関数として用いられる⁹⁾。SSIM は以下の式で表される。

$$SSIM(x, \hat{x}) = \frac{2\mu_x\mu_{\hat{x}} + C_1}{\mu_x^2 + \mu_{\hat{x}}^2 + C_1} \times \frac{2\sigma_{x\hat{x}} + C_2}{\sigma_x^2 + \sigma_{\hat{x}}^2 + C_2} \quad \dots\dots(3)$$

ここで、 $\mu_x, \mu_{\hat{x}}$ はそれぞれ実画像 x 、出力した画像 $\hat{x}$ の平均画素値、 $\sigma_x, \sigma_{\hat{x}}$ はそれぞれ x と $\hat{x}$ の画素値の標準偏差、 $\sigma_{x\hat{x}}$ は共分散、 C_1, C_2 は定数であり、本手法では $C_1 = 0.01, C_2 = 0.03$ とした。SSIM は両画像間の構造が一致すれば1となるため、目的関数 L_{SSIM} は 式(4)に変換し、Adversarial loss と同様に、最小値を得ることを目的とする。

$$L_{\text{SSIM}} = 1 - SSIM(x, \hat{x}) \quad \dots\dots(4)$$

3. 結果

提案手法の有効性を確認するために、評価実験を行う。シングルショット画像に対してノイズ除去を行った画像と、120枚加算平均画像との画質を比較する。

120枚加算平均画像と比較する対象として、画像のノイズ除去として一般的な画像処理フィルタを3手法、Deep learning によるノイズ除去を2手法用いる。

- ・ ガウシアンフィルタ：近傍の距離に基づく重み付き平均
- ・ メディアンフィルタ：近傍の中央値
- ・ Non-local Means フィルタ¹⁰⁾：近傍の局所領域の類似度に基づく重み付き平均 (NL Means フィルタ)
- ・ CNN：SSIM loss
- ・ 提案手法：CNN (Adversarial loss + SSIM loss)

3.1 客観的評価

本実験では、120枚加算平均画像と各手法のノイズ除去結果の評価値を求める。評価値として、画質評価として用いられる PSNR (Peak signal-to-noise ratio) と2.2で述べた SSIM の2つを用いる。PSNR は画像の劣化を指標化したものであり、数値は高いほど良い。SSIM は構造的な類似性の評価を指標化したものであり、数値は1に近いほど良い。

表 1 120 枚加算平均画像とそれぞれの手法の PSNR/SSIM

	シングル ショット画像	ガウシアン フィルタ	メディアン フィルタ	NL Means フィルタ	CNN	提案手法
PSNR [dB]	20.78	27.91	23.98	28.78	32.44	31.35
SSIM	0.22	0.63	0.47	0.81	0.89	0.89

表 2 各手法の DMOS 値

	シングル ショット画像	ガウシアン フィルタ	メディアン フィルタ	NL Means フィルタ	CNN	提案手法	120 枚 加算平均画像
DMOS 値	1.70	1.84	1.62	1.90	3.77	4.31	4.79

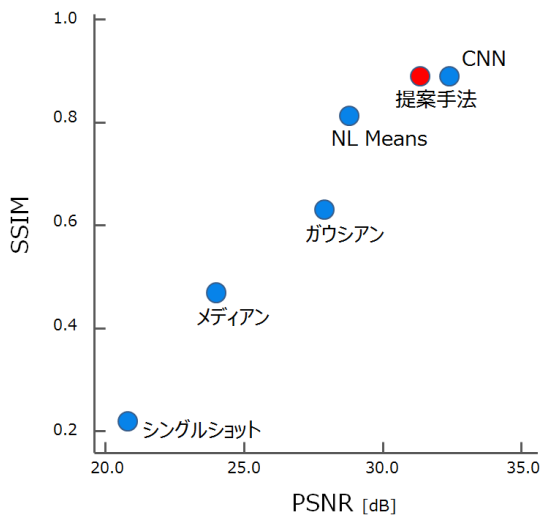


図 5 120 枚加算平均画像に対する各手法の PSNR/SSIM
120 枚加算平均画像に対し、CNN と提案手法が共に PSNR, SSIM が高い値である。

図 4 に各手法のノイズ除去結果、表 1 に図 4 の各画像に対する PSNR と SSIM を示す。また、図 5 に各手法の PSNR と SSIM の分布を示す。

客観的評価の結果、提案手法を含む Deep learning によるノイズ除去画像は PSNR, SSIM 共に高い数値を得ることができた。よって、シングルショットの OCT 画像から 120 枚加算相当の画像が出力できていると考えられる。なお、客観的評価では、CNN は有効であるが、loss による違いはないことが確認できる。

3.2 主観的評価

主観的評価として、ITU-T 勧告 P.800 に規定される平均オピニオン評点 (Mean Opinion Score, MOS) がある。MOS は、音声、画像、映像、視聴覚コンテンツなど様々な分野の品質を定量化するための指標として用いられ

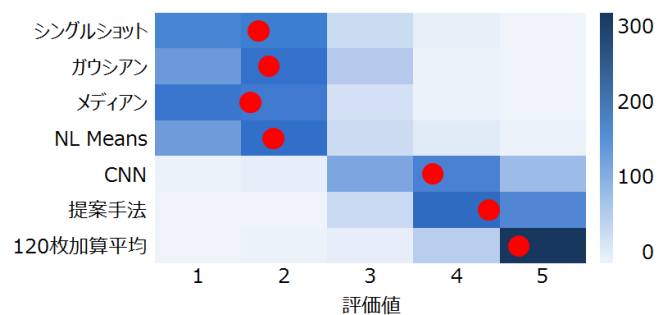


図 6 各手法の評価値分布と DMOS 値

各手法における全評定 (10 枚の画像, 37 名の評定者) の分布であり、色が濃い評価ほど評定数が多いことを示す。図中の赤丸は DMOS 値を示す。

ている¹¹⁾。本評価では、120 枚加算平均画像を参照画像とし、参照画像と評価対象画像にどれだけ違いが認められるかを評価する Degradation MOS (DMOS) を用いる。

評価は、評定者 37 名に対し、評価対象画像とそれに対応する 120 枚加算画像を提示し、その違いを評価させる。評価は 5 段階で、画像中のノイズの程度、網膜層の構造の違いを総合的な判断から、「ほとんど違いが認められない (5)」「違いが認められるが気にならない (4)」「違いが認められ、わずかに気になる (3)」「違いが認められ、気になる (2)」「違いが認められ、非常に気になる (1)」とした。なお、評価対象には必ず最高点の画像が存在するように、参照画像である 120 枚加算画像を含めた。さらに、評定者には評価対象に 7 パターンの画像が含まれること、120 枚加算画像が含まれることを伝えず、70 枚の画像をランダムに与えた。評価対象画像は、提案手法と 5 つの比較手法によって得られた画像に加え、シングルショット画像と 120 枚加算平均画像の 7 パターンとし、これら 10 組 (いずれも

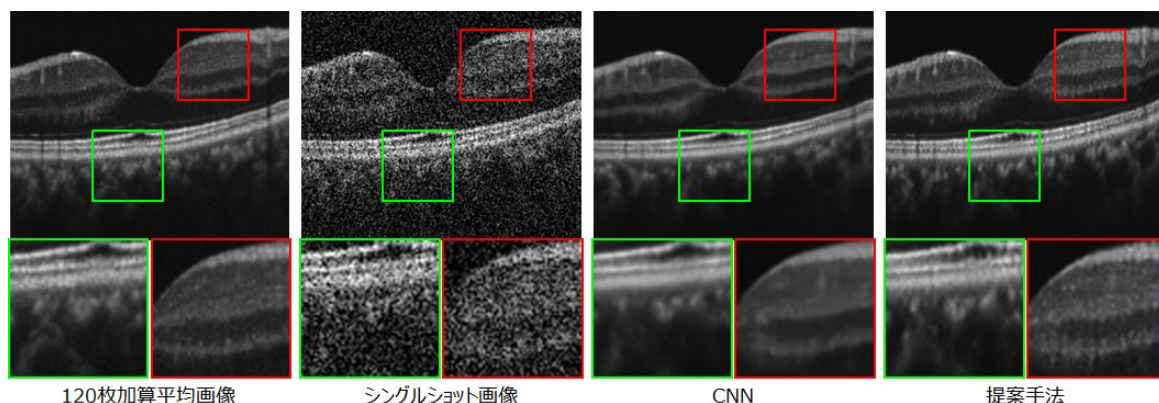


図 7 提案手法と CNN による同一部位の OCT 画像

GAN を用いた提案手法によるノイズ除去画像は、網膜の境界や脈絡膜領域で鮮明に出力されており、120 枚加算平均画像により近い画像である。

学習に用いていない検証用データ), 計 70 枚を用意した。

各評価者によって得られた評価値は(5)式を用いて DMOS 値として算出する。

$$DMOS = \frac{1}{n} \sum_{t=1}^n \alpha_t \quad \dots\dots(5)$$

ここで、 n は評価者数、 α_t は評価者 t の評価値である。図 6、および表 2 に評価結果を示す。

主観的評価の結果、提案手法が 120 枚加算平均画像に次ぐ高い DMOS 値となった。また、提案手法とその次に DMOS 値が高い CNN の結果について符号検定を行ったところ、 $p=5.18 \times 10^{-10}$ であり、有意差があることが確認できた。

以上のことから、CNN と提案手法の評価値には差が認められ、GAN に用いられる Adversarial loss が 120 枚加算平均画像に近い画像の出力に貢献していると考えられる。参考として、図 7 に GAN を用いた提案手法によるノイズ除去画像と、CNN によるノイズ除去画像の同一領域の拡大画像を示す。

4. 考按

本稿では、Adversarial loss を取り入れた Deep learning を用いたシングルショット OCT 画像からのノイズ除去手法を提案した。客観的評価では、提案手法と CNN 共に高い評価値を得ることで、OCT 画像ノイズ除去に対する CNN の有効性を確認した。さらに主観評価により Adversarial loss を用いた本手法の有効性を確認した。

提案手法を加算平均処理の代わりに用いることで、被験者の負担の軽減や、固視状態の悪い被検者の観察に貢献できると考えられる。

本実験では、学習データとして正常眼のみを用いており疾患眼の画像は学習に用いていない。図 8 に本実験で用いた正常眼のみを学習したモデルを疾患眼に適用した例を示す。提案手法によるノイズ除去画像は網膜の構造が崩れた部位で多少ボケがみられる。疾患眼は正常眼とは異なる構造を含むため、このような構造に対応するには、様々な疾患眼データを学習データに加える必要があると考えられる。また、実臨床に適用する際に重要となるのは、病的所見が判別できるかどうかである。図 8 より、本手法の結果はシングルショット画像、120 枚加算平均画像と比べ網膜構造の大きな違いは見られず、実臨床への悪影響もないことが期待される。よって今後は、疾患眼を含むデータセットを作成し、実臨床への影響を評価することが課題となる。

また、提案手法はシングルショット画像、すなわち 1 回の B スキャンの画像からノイズを除去した画像を予測するため、入力したシングルショット画像に写っていない構造を予測することはできない。図 9 に示すように、赤血球の反射はスキャンした時間によって写る場合と写らない場合があるため、シングルショット画像取得時に赤血球反射が写っていない場合は、画像として出力することができない。本手法を加算平均処理の代わりとして使う場合は、この点に注意する必要がある。

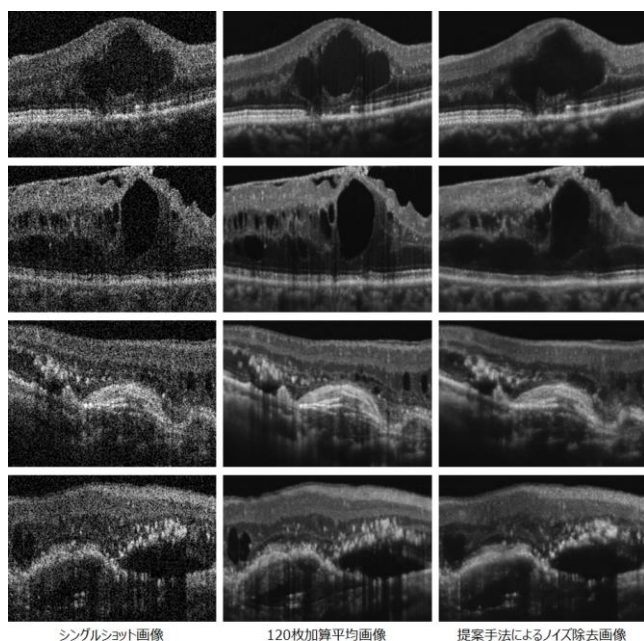


図 8 疾患眼に対するノイズ除去

正常眼のみ学習させたネットワークに、学習に用いていない疾患画像を入力し、ノイズ除去を実施。

謝 辞

研究遂行にあたり臨床データのご提供を賜りました鹿児島大学医学部 眼科学教室、医療法人 明星会 園田病院の園田祥三先生に心より感謝いたします。

文 献

- 1) 板谷正紀：眼底用光干渉断層計の進歩．MEDICAL PHOTONICS No.7, 58-64, 2011.
- 2) Sander B., Larsen M., et al.: Enhanced optical coherence tomography imaging by multiple scan averaging. British Journal of Ophthalmology 89 (2), 207-212, 2005.
- 3) Sripad K. D., Giridhar S., et al.: A Deep Learning Approach to Denoise Optical Coherence Tomography Images of the Optic Nerve Head. arXiv preprint arXiv:1809.10589, 2018.
- 4) Isola P., Zhu J. Y., et al.: Image-to-image translation with conditional adversarial networks. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
- 5) Kazeminia S., Baur C., et al.: GANs for Medical Image Analysis. arXiv preprint arXiv:1809.06222, 2018.
- 6) Yi X., Walia E., et al.: Generative Adversarial Network in Medical Imaging: A Review. arXiv preprint arXiv:1809.07294, 2018.
- 7) Ronneberger O., Fischer P., et al.: U-Net: Convolutional networks for biomedical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015.
- 8) Wang A., Bovik A. C., et al.: Image quality assessment: from error visibility to structural similarity. IEEE

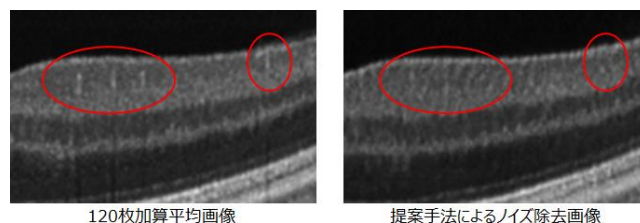


図 9 シングルショット画像からは生成できない情報の例

Deep Learning は、入力する画像に存在しない情報を予測することはできないため、シングルショット画像に赤血球反射が写っていない場合は、出力画像に反映することができない。

Transactions on Image Processing Vol.13, 600-612, 2004.

- 9) Hang Z., Orazio G., et al.: Loss Functions for Image Restoration With Neural Networks. IEEE Transactions on Computational Imaging Vol.3, 47-57, 2017.
- 10) Buades A., Coll B., et al.: A non-local algorithm for image denoising. IEEE Computer Vision and Pattern Recognition (CVPR) Vol.2, 60-65, 2005.
- 11) Streijl C. R., Winkler S, et al.: Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives. Multimedia Systems Vol. 22(2), 213-227, 2016.