Joint Optimization of Convolutional Neural Network and Image Preprocessing Selection for Embryo Grade Prediction in *In Vitro* Fertilization*

Kento Uchida**[1], Shota Saito**[1,3], Panca Dewi Pamungkasari**[1,2], Yusei Kawai[1], Ita Fauzia Hanoum[4], Filbert H. Juwono[5], and Shinichi Shirakawa[1]

[1] Graduate School of Environment and Information Sciences, Yokohama National University, Yokohama, Japan
[2] Department of Information and Communication Technology, Universitas Nasional, Jakarta, Indonesia
[3] SkillUp AI Co., Ltd., Tokyo, Japan
[4] Infertility Laboratory of Permata Hati Program, RSUP DR Sardjito, Indonesia
[5] Department of Electrical and Computer Engineering, Curtin University Malaysia, Sarawak, Malaysia

Abstract. Convolutional neural networks (CNNs) are a standard tool for image recognition. To improve the performance of CNNs, it is important to design not only the network architecture but also the preprocessing of the input image. Extracting or enhancing the meaningful features of the input image in the preprocessing stage can help to improve the CNN performance. In this paper, we focus on the use of the well-known image processing filters, such as the edge extraction and denoising, and add the preprocessed images to the input of CNNs. As the optimal filter selection depends on dataset, we develop a joint optimization method of CNN and image processing filter selection. We represent the image processing filter selection by a binary vector and introduce the probability distribution of the vector. To derive the gradient-based optimization algorithm, we compute the gradients of weight and distribution parameters on the expected loss under the distribution. The proposed method is applied to an embryo grading task for *in vitro* fertilization, where the embryo grade is assigned based on the morphological criterion. The experimental result shows that the proposed method succeeds to reduce the test error by more than 8% compared with the naive CNN models.

Keywords: Convolutional Neural Network · Image Processing Filter · Embryo Grading · *In Vitro* Fertilization

* This is an author version of the paper accepted to the 14th International Symposium on Visual Computing (ISVC 2019). The final authenticated version is available online at `https://doi.org/10.1007/978-3-030-33723-0_2`.

** Equal contribution

1 Introduction

Convolutional neural networks (CNNs) continue to update the state-of-the-art performance of image classification and detection. The performance improvement of CNNs is not only caused by the architecture design but also preprocessing of input images. Traditional image processing filters, such as denoising and edge enhancement filters, help to remove unuseful noise or extract important features in input images. For instance, Calderon et al. [4] analyze the impact of denoising, contrast, and edge enhancement as a preprocessing for CNN, and show that such preprocessing has the potential to improve the performance of CNN in X-ray image dataset.

Adequate preprocessing should be selected depending on the datasets to exploit the potential of image preprocessing for CNNs. We focus on the problem to select the adequate image processing filters as preprocessing from candidates prepared in advance. In this paper, we develop a method for optimizing the image processing filter selection during a single CNN training. In our method, a binary vector represents the selection of image processing filters, i.e., each bit indicates that the corresponding filter is selected or not. It is difficult to optimize the binary vector by gradient method because it is discrete. To handle this difficulty, we consider the multivariate Bernoulli distribution as the underlying distribution of the binary vector and introduce an alternative differentiable objective function parameterized by the distribution parameters and the weights of a neural network. This relaxation technique is called *stochastic relaxation*, which is used in the one-shot neural architecture search [1,13] and embedded feature selection [14]. Instead of directly optimizing the binary vector, we optimize the distribution parameters jointly with the weights. Figure 1 shows the conceptual image of the proposed method. Since the proposed method requires the model training only once, the increase of computational time is not significant.

We believe that the image processing filters, such as edge detection and denoising, are useful for medical images rather than for general object images, because a small difference in the edge or shape may be important in medical images. To verify the effect of our selection scheme of image processing filters, we apply the proposed method to the prediction of embryo grade for *in vitro* fertilization (IVF) using day 3 embryo images. The experimental result shows that the proposed method can reduce the test error by more than 8% compared with the naive CNN model without image processing filters.

2 Joint Optimization of CNN and Image Preprocessing Filter Selection

We consider the general problem setting on supervised learning with a given training image dataset $\mathcal{D} = \{X, Y\}$. Let $X = \{x_1, \ldots, x_N\}$ be the set of input images, and $Y = \{y_1, \ldots, y_N\}$ be the set of corresponding target variables. The goal of a machine learning algorithm is to find a better-generalized model $\phi : x \mapsto y$ that works well on unseen (test) data (x, y).

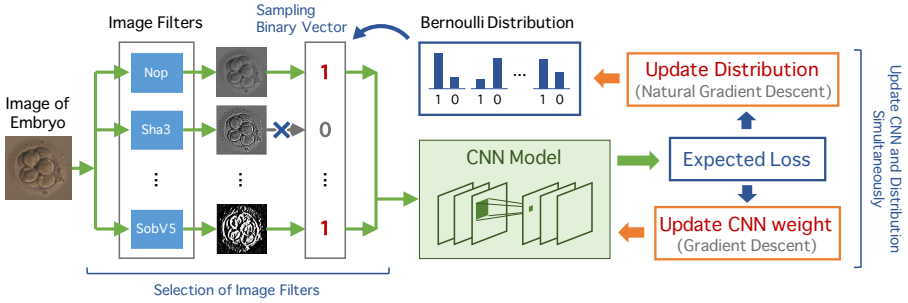

Fig. 1. Conceptual illustration of the proposed method.

In this paper, we consider a CNN model with image preprocessing denoted by $\phi(W, M)$ parameterized by the weight vector W and binary vector $M \in \mathcal{M} = \{0, 1\}^d$, where M represents the filter selection. Each bit in M determines whether or not the preprocessed image by the corresponding filter appears in the input images for CNN.

Specifically, we prepare the image processing filters, $\psi_1, \ldots, \psi_d$, that transform an original image x into the filtered images $\bar{x} = \{\psi_1(x), \ldots, \psi_d(x)\}$. These filters are expected to extract or enhance the useful features of the original image. The filtered images are concatenated to the channel direction and used as the input of CNN. We determine whether each filtered image is included in the input of CNN by the binary vector. Therefore, the input of CNN can be obtained by $\{m_1\psi_1(x), \ldots, m_d\psi_d(x)\}$, where $m_i \in \{0, 1\}$ indicates i-th element of M. As the useful image processing filters depend on datasets and the unmeaningful filters may lead the performance deterioration, we develop the algorithm that finds the better binary vector in a single CNN training.

Given a loss function $\mathcal{L}(W, M)$ to be minimized, such as the cross-entropy, the gradient with respect to (w.r.t.) M cannot be obtained because M is discrete variable. We transform the original objective into the differentiable one by using the probability distribution of M. Let us consider the multivariate Bernoulli distribution $p_\theta(M) = \prod_{i=1}^{d} \theta_i^{m_i}(1 - \theta_i)^{1-m_i}$, where $\theta \in \Theta \subseteq [0, 1]^d$ is the distribution parameter. The value θ_i is the probability that m_i becomes 1 and can be regarded as the selection probability of the i-th filter. By taking the expectation of $\mathcal{L}(W, M)$ under the distribution, we get an alternative objective function for W and θ: $\mathcal{G}(W, \theta) = \mathbb{E}_{p_\theta}[\mathcal{L}(W, M)]$. This transformation of objective function is called *stochastic relaxation*. Different from the original objective function $\mathcal{L}$, the transformed objective function $\mathcal{G}$ is differentiable w.r.t. both W and θ.

We adopt the optimization algorithm proposed in [13,14] to minimize the expected loss $\mathcal{G}(W, \theta)$. The vanilla (Euclidian) gradient w.r.t. W and the natural gradient w.r.t. θ are given by $\nabla_W \mathcal{G}(W, \theta) = \sum_{M \in \mathcal{M}} \nabla_W \mathcal{L}(W, M) p_\theta(M)$ and $\tilde{\nabla}_\theta \mathcal{G}(W, \theta) = \sum_{M \in \mathcal{M}} \mathcal{L}(W, M) \tilde{\nabla}_\theta \ln p_\theta(M) p_\theta(M)$, respectively, where $\tilde{\nabla}_\theta = F(\theta)^{-1}\nabla_\theta$ is the so-called natural gradient [3] that is the steepest direction of θ w.r.t. the Kullback-Leibler divergence, and $F(\theta)$ is the Fisher information ma-

trix of p_θ. The training of the model ϕ is done by iteratively updating both W and θ to the above gradient direction.

In practice, these gradients are approximated by Monte-Carlo using λ samples, $M_1, \ldots, M_\lambda$, drawn from p_θ. Also, the loss function is approximated using N mini-batch data samples $\mathcal{Z}$ as $\mathcal{L}(W, M) \approx \bar{\mathcal{L}}(W, M; \mathcal{Z}) = \frac{1}{N} \sum_{z \in \mathcal{Z}} l(z, W, M)$, where $l(z, W, M)$ represents the loss of a datum. Consequently, the gradient w.r.t. W is given by

$$\nabla_W \mathcal{G}(W, \theta) \approx \frac{1}{\lambda} \sum_{i=1}^{\lambda} \nabla_W \bar{\mathcal{L}}(W, M_i; \mathcal{Z}) \ . \tag{1}$$

The gradient $\nabla_W \bar{\mathcal{L}}(W, M_i; \mathcal{Z})$ can be computed using back-propagation and we can use any stochastic gradient descent (SGD) method for optimizing W.

Since we consider the Bernoulli distribution as p_θ, the natural gradient of the log-likelihood can be analytically obtained as $\tilde{\nabla} \ln p_\theta(M) = M - \theta$. Then, we can get the approximation of the natural gradient w.r.t. θ as $\tilde{\nabla}_\theta \mathcal{G}(W, \theta) \approx \frac{1}{\lambda} \sum_{i=1}^{\lambda} \bar{\mathcal{L}}(W, M_i; \mathcal{Z})(M_i - \theta)$. For the natural gradient estimate, we transform the loss value into the ranking-based utility as was done in [13,14] as follows: $u_i = 1$ for best $\lceil \lambda/4 \rceil$ samples, $u_i = -1$ for worst $\lceil \lambda/4 \rceil$ samples, otherwise $u_i = 0$. The ranking-based utility transformation makes the algorithm invariant to the order preserving transformation of $\mathcal{L}$. With this utility transformation, the θ update becomes

$$\theta^{(t+1)} = \theta^{(t)} + \frac{\eta_\theta}{\lambda} \sum_{i=1}^{\lambda} u_i (M_i - \theta^{(t)}) \ , \tag{2}$$

where η_θ is the learning rate for θ. We note that the minimization problem is turned into the maximization problem due to the utility transforms. In the algorithm implementation, we restrict the range of θ_i within $[1/d, 1 - 1/d]$ to keep the possibility of generating any binary vector. The training procedure is summarized in Algorithm 1.

After the model training, we choose the most likely binary vector to predict a new data as $\hat{M} = \mathrm{argmax}_M \, p_\theta(M)$ such that $m_i = 1$ if $\theta_i \geq 0.5$, otherwise $m_i = 0$. Then, we predict new data using the image processing filters decided by $\hat{M}$.

3 Embryo Grading Task and Dataset

IVF is the most common fertility treatment which assists a woman for getting pregnancy with genetic technology. The low success rate of IVF has become a major issue to date. According to [6], embryo grading is responsible for around one-third of implantation failures. Hence, the determination of the morphological embryo grading for the base of the selection is very important. Currently, the embryo selection is done through a manual grading by embryologist using a microscope. The manual selection makes the process dependent on the embryologist's skills and experiences. Therefore, embryo grading is subjective and the

Algorithm 1: The training procedure of the proposed method.

Input: Training data $\mathcal{D}$ and hyperparameters $\{\lambda, \eta_\theta\}$
Output: Optimized parameter of W and θ

begin
  Initialize the connection weights of CNN and Bernoulli distribution parameter as $W^{(0)}$ and $\theta^{(0)}$
  $t \leftarrow 0$
  while *not stopping criterion is satisfied* **do**
    Get N mini-batch samples from $\mathcal{D}$
    Sample $M_1, \ldots, M_\lambda$ from $p_{\theta^{(t)}}$
    Compute the loss $\bar{\mathcal{L}}(W, M_i)$ for each sample
    Update the distribution parameter to $\theta^{(t+1)}$ using (2)
    Force $\theta^{(t+1)} \in [1/d, 1 - 1/d]^d$
    Update the connection weights to $W^{(t+1)}$ using (1) by any SGD
    $t \leftarrow t + 1$

fate of an embryo is determined only based on restricted information and observation. The machine learning techniques have been applied to embryo images to automate and improve the embryo grading process. For instance, the number of cells in day 3 embryo, one of the important criterion to determine the embryo grade, was predicted using features extracted from embryo image by conditional random fields [11] or Adaboost [19]. Khan et al. [8] showed that CNN could improve the performance of cell counting in embryo image. Also, CNN was used to predict the embryo grading from the fifth day's embryo image directly [5,9].

In this paper, we use the day 3 embryo images that were collected from 2016 to 2018 at the Infertility Clinic in Indonesia. The day 3 embryo grading is more crucial than the day 5 embryo grading since it makes the decision whether the embryo will be transferred to uterus or cultured until the blastocyst phase [2,12]. The grade of the embryo was categorized by an embryologist based on the standard morphological criterion introduced by [18]. Due to the limitation of the dataset, the grades for embryos were adjusted to be 3 groups; they could be either excellent, moderate, or poor. Table 1 shows the correspondence of our grading scheme with the grading of [18] and the morphological criterion. The dataset consisted of 254, 599, and 533 images for excellent embryos, moderate embryos, and poor embryos, respectively. The total number of images was 1386, which was collected from 238 patients, and each patient had about 4 embryos on the 3rd day. Figure 2 shows the examples of the embryo for each grade. The task is to classify the embryo images into three categories (grades). In the experiment, the images were resized to 128 $\times$ 128 and converted to grayscale. Then, we randomly assigned 80 % of all images as the training dataset, and 20 % as the test dataset.

Table 1. Correspondence of the modified grading applied to our dataset with the grade introduced by [18] and morphological criterion for the embryo on the 3rd day.

Grade [18]	Modified Grade	Number of Cells	Blastomere	Fragmentation
1	Excellent	7-8 cells	Similar	less than 5%
2	Moderate	7-8 cells	Similar	$5-15\%$
3	Poor	6-8 cells	Fairly similar	$15-25\%$
4	Poor	4-6 cells	Fairly similar	$25-30\%$
5	Poor	4-5 cells	Not Similar	more than 30%

Fig. 2. Examples of embryo image in excellent (left), moderate (center), and poor (right) grades.

4 Experiment and Result

4.1 Experimental Setting

The CNN architecture used in the experiment is based on the VGG-Net [15] consisting of 16 convolutional layers with 3×3 kernel. Different from the original VGG-Net, we add the batch normalization in the convolutional and fully-connected layers and remove the dropout [16] in the fully-connected layers. In addition, we insert the global average pooling [10] before the first fully-connected layer. As our embryo grading task is classification, the loss function is the softmax cross-entropy.

We set the mini-batch size to 64 and the maximum number of epochs to 3,000 (about 54,000 training iterations). The weight parameters are initialized by He's initialization [7] and optimized by using Nesterov's accelerated stochastic gradient method [17] with a momentum of 0.9. We also use the weight decay of 10^{-4} for the training of the weight parameters. The learning rate for the weights is initialized by 0.1 and divided by 10 at 1/2 and 3/4 of the maximum number of epochs. These parameter settings are based on [13].

We prepare 15 image processing filters including no operation (Nop) listed in Table 2. The sample size of the binary vector in each iteration is $\lambda = 2$, and the learning rate for distribution parameters is set to $\eta_\theta = 1/15$. The distribution parameters θ are initialized by 0.5. Each preprocessed image by the image processing filters is standardized, i.e., each pixel value is subtracted by the mean

Table 2. Image processing filters used in the experiment.

Filter	Abbreviation		
	–	(3x3 kernel)	(5x5 kernel)
Original image	Nop	–	–
Blur filter	–	Blur3	Blur5
Gaussian filter	–	Gau3	Gau5
Sharpening filter	–	Sha3	Sha5
Laplacian filter	–	Lap3	Lap5
Vertical-Sobel filter	–	SobV3	SobV5
Horizontal-Sobel filter	–	SobH3	SobH5
Median filter	–	Med3	Med5

pixel value and divided by the standard deviation. Also, we apply the standard data augmentation, i.e., shifting, flipping, and rotation, to the filtered images.

To investigate the effect of the proposed method, we compare the following four models:

1. **No Filter**: the model without image preprocessing filters (naive CNN)
2. **All Filters**: the model using all preprocessed images as inputs
3. **Random**: the model with fixed filter selection probability of 0.5 (i.e., our method without θ optimization)
4. **Joint Optimization (ours)**: the model that are jointly optimized CNN and the image preprocessing filter selection

We report the experimental results based on five trials with different random seeds. For the compared models, No Filter, All Filters, and Random, we set the mini-batch size to 128. The numbers of data samples used to the parameters update become equal among all models by setting such batch size because our method uses the doubled number of data samples ($\lambda = 2$) in each iteration.

4.2 Result and Discussion

Figure 3 shows the transitions of the median value and the inter-quartile range of test error for each model. Our method can reduce the test error than the other models after about 5000 iterations. The values of median (lower and upper quartile) of the test error at the final iteration are as follows; No Filter: 39.57% (37.05%, 39.93%), All Filters: 38.13% (38.13%, 42.45%), Random: 46.04% (44.24%, 50.72%), and Joint Optimization (ours): 28.78% (26.62%, 28.78%). We observe that our method significantly reduces the test error at the final iteration by more than 8 % compared to No Filter and All Filters. This result implies that the redundant image preprocessing has the potential to lead a adverse effect, but selecting appropriate filters can improve the performance of CNN. The result of Random significantly worse than other models, suggesting that the optimization algorithm of our method can work well. We note that the total training time of all models is more or less the same. Thus, the proposed

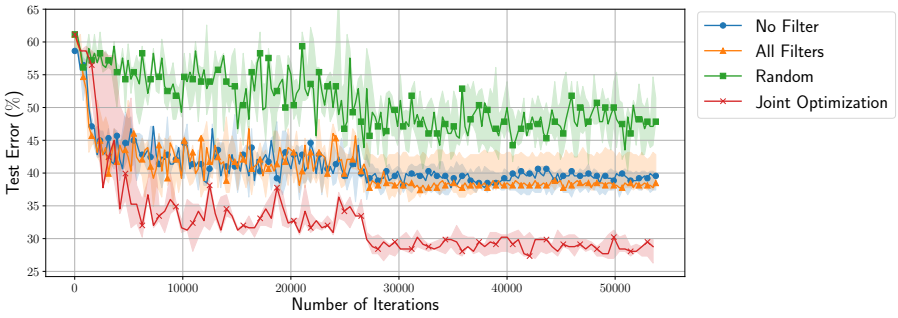

Fig. 3. Transitions of test error for each model.

method can improve the prediction accuracy without additional computational cost.

Figure 4 depicts the confusion matrices of test data for each model. We can see that the recall obtained by our method is higher than others for moderate and poor grades. For the excellent grade, the accuracy of all models is not so high, and the excellent embryo tends to be predicted incorrectly as the moderate grade. According to Table 1, the difference between the excellent grade and moderate grade is only the amount of fragmentation. We believe that this poor accuracy for excellent embryo data is because the number of data graded as excellent is not sufficient for learning such differences. We emphasize that the ratio of the worst misclassification, where the excellent embryo is predicted as poor, or the inverse case, is very low in our method. From the viewpoint of IVF, this result is encouraging as both of the excellent and moderate embryos can be transferred to the uterus. Moreover, distinguishing poor embryo is the most important task because the worst situation is transferring the poor embryo to the uterus. This improvement suggests that the selected image processing filters can extract useful features for accurate grade prediction.

Figure 5 shows the preprocessed images and their selection probabilities at the final iteration. We observe that the selection probability of Nop is not high, implying that the preprocessed images are useful in embryo grade prediction rather than the original image. The filters whose selection probability is more than 0.8 are Sha3, Sha5, SobV3, SobV5, SobH3, SobH5, Med3, and Med5. Regarding the criterion of embryologist in Table 1, Med3 and Med5 contribute to eliminating the fragmentation, and Sha3 and Sha5 conversely emphasize them. The edge detection filters (SobV3, SobV5, SobH3, and SobH5) are expected to help to count the embryo and to determine the similarity of blastomere.

5 Conclusion

We have proposed the method that jointly optimizes CNN and the image processing filter selection. We introduced the Bernoulli distribution as a law of

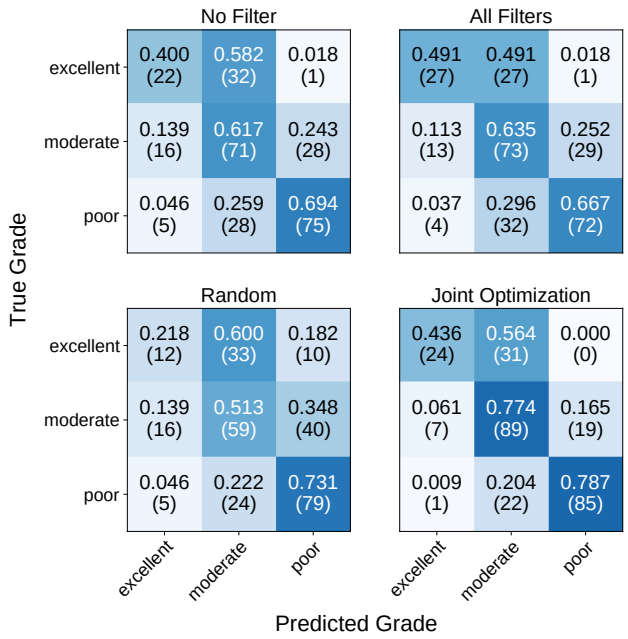

Fig. 4. Confusion matrices of test data for each model. The values in each cell indicate the ratio of data samples classified into each category, and the values in parentheses indicate the number of data samples. We report the confusion matrix of the trial that recorded the median of test error for each model.

binary vector for filter selection. The distribution parameter and the connection weight of CNN were jointly updated to the gradient directions. We applied the proposed method to the embryo grading prediction for IVF, where the labels of the dataset were given by an embryologist based on the morphological criteria. We have shown that the test error of the proposed method can be reduced by more than 8 % compared to naive CNN.

The different types of image processing filters are often applied several times to extract and enhance meaningful features. In our method, each image processing filter is employed only once to the original image. Optimizing the combination of image processing filter sequences is one of the future works. In the embryo image dataset, we consider that the prediction of the excellent grade directly is difficult because its criteria are similar to those of moderate grade. Therefore, constructing the CNN that predicts the criteria for embryo grading, i.e., cell number, the similarity of blastomere and amount of fragmentation, may contribute to the improvement of the prediction accuracy for embryo grading.

In the experiment of this paper, most of the prepared filters were linear filters and can be represented by a convolution operation in CNNs. Nevertheless, we confirmed that the proposed method is useful in our embryo dataset. Although the evaluation of our method on other datasets, such as large scale dataset,

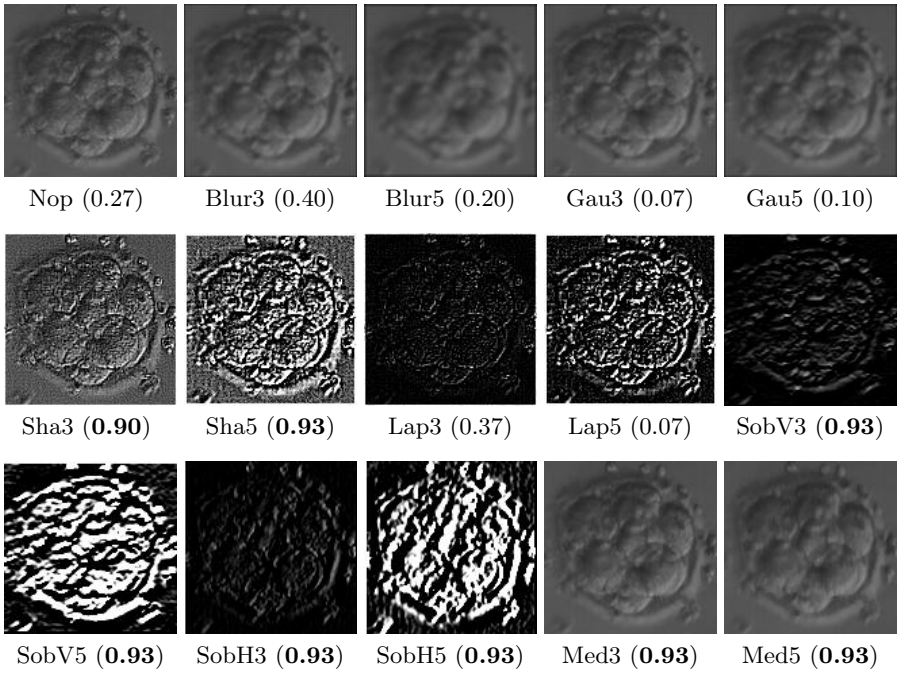

Fig. 5. Preprocessed images of the embryo by the image processing filters. The values in the parentheses are the selection probability of each filter, where the bold font means the selection probability is higher than 0.8. The result is the trial that recorded the median of test error.

is another important future work. In addition, as our method can adopt non-linear and non-differentiable image preprocessing, adding such non-linear filters is promising future work.

References

1. Akimoto, Y., Shirakawa, S., Yoshinari, N., Uchida, K., Saito, S., Nishida, K.: Adaptive Stochastic Natural Gradient Method for One-Shot Neural Architecture Search. In: Proceedings of the 36th International Conference on Machine Learning (ICML). vol. 97, pp. 171–180 (2019)
2. Alfaraj, S., Alzaher, F., Alshwaiaer, S., Ahmed, A.: Pregnancy Outcome of Day 3 versus Day 5 Embryo Transfer: A Retrospective Analysis. Asian Pacific Journal of Reproduction **6**(2), 89–92 (2017). https://doi.org/10.12980/apjr.6.20170208
3. Amari, S.: Natural Gradient Works Efficiently in Learning. Neural Computation **10**(2), 251–276 (1998)
4. Calderon, S., Fallas, F., Zumbado, M., Tyrrell, P.N., Stark, H., Emersic, Z., Meden, B., Solis, M.: Assessing the Impact of the Deceived Non Local Means Filter as a Preprocessing Stage in a Convolutional Neural Network Based Approach for Age Estimation Using Digital Hand X-Ray Images. In: 25th IEEE

International Conference on Image Processing (ICIP). pp. 1752–1756 (2018). https://doi.org/10.1109/ICIP.2018.8451191
5. Chen, T.J., Zheng, W.L., Liu, C.H., Huang, I., Lai, H.H., Liu, M.: Using Deep Learning with Large Dataset of Microscope Images to Develop an Automated Embryo Grading System. Fertility & Reproduction **01**(01), 51–56 (2019). https://doi.org/10.1142/S2661318219500051
6. Craciunas, L., Gallos, I., Chu, J., Bourne, T., Quenby, S., Brosens, J.J., Coomarasamy, A.: Conventional and Modern Markers of Endometrial Receptivity: a Systematic Review and Meta-Analysis. Human Reproduction Update **25**(2), 202–223 (2019). https://doi.org/10.1093/humupd/dmy044
7. He, K., Zhang, X., Ren, S., Sun, J.: Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In: IEEE International Conference on Computer Vision (ICCV). pp. 1026–1034 (2015). https://doi.org/10.1109/ICCV.2015.123
8. Khan, A., Gould, S., Salzmann, M.: Deep Convolutional Neural Networks for Human Embryonic Cell Counting. In: Computer Vision – ECCV 2016 Workshops. pp. 339–348 (2016)
9. Khosravi, P., Kazemi, E., Zhan, Q., E. Malmsten, J., Toschi, M., Zisimopoulos, P., Sigaras, A., Lavery, S., A. D. Cooper, L., Hickman, C., Meseguer, M., Rosenwaks, Z., Elemento, O., Zaninovic, N., Hajirasouliha, I.: Deep Learning Enables Robust Assessment and Selection of Human Blastocysts after in Vitro Fertilization. Nature Partner Journals Digital Medicine **02**(21) (2019). https://doi.org/10.1038/s41746-019-0096-y
10. Lin, M., Chen, Q., Yan, S.: Network in Network. In: International Conference on Learning Representations (ICLR) (2014)
11. Moussavi, F., Wang, Y., Lorenzen, P., Oakley, J., Russakoff, D., Gould, S.: A Unified Graphical Models Framework for Automated Human Embryo Tracking in Time Lapse Microscopy. In: IEEE 11th International Symposium on Biomedical Imaging. pp. 314–320 (2014)
12. Racowsky, C., Jackson, K.V., Cekleniak, N.A., Fox, J.H., Hornstein, M.D., Ginsburg, E.S.: The Number of Eight-cell Embryos is a Key Determinant for Selecting Day 3 or Day 5 Transfer. Fertility and Sterility **73**(3), 558–564 (2000). https://doi.org/10.1016/S0015-0282(99)00565-8
13. Shirakawa, S., Iwata, Y., Akimoto, Y.: Dynamic Optimization of Neural Network Structures Using Probabilistic Modeling. In: Thirty-Second AAAI Conference on Artificial Intelligence (AAAI). pp. 4074–4082 (2018)
14. Shota, S., Shirakawa, S., Akimoto, Y.: Embedded Feature Selection Using Probabilistic Model-based Optimization. In: Proceedings of the Genetic and Evolutionary Computation Conference Companion (GECCO). pp. 1922–1925 (2018). https://doi.org/10.1145/3205651.3208227
15. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. In: International Conference on Learning Representations (ICLR) (2015)
16. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research **15**, 1929–1958 (2014)
17. Sutskever, I., Martens, J., Dahl, G., Hinton, G.: On the importance of initialization and momentum in deep learning. In: Proceedings of the 30th International Conference on Machine Learning (ICML). vol. 28, pp. 1139–1147 (2013)

18. Veeck, L.: An Atlas of Human Gametes and Conceptuses: An Illustrated Reference for Assisted Reproductive Technology. New York: The Parthenon Publishing Group (1999)
19. Wang, Y., Moussavi, F., Lorenzen, P.: Automated Embryo Stage Classification in Time-Lapse Microscopy Video of Early Human Embryo Development. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013. pp. 460–467 (2013)