Adaptable image quality assessment using meta-reinforcement learning of task amenability

Shaheer U. Saeed[1], Yunguan Fu[1,2], Vasilis Stavrinides[3,4], Zachary M. C. Baum[1], Qianye Yang[1], Mirabela Rusu[5], Richard E. Fan[6], Geoffrey A. Sonn[5,6], J. Alison Noble[7], Dean C. Barratt[1], and Yipeng Hu[1,7]

[1] Centre for Medical Image Computing, Wellcome/EPSRC Centre for Interventional & Surgical Sciences, and Department of Medical Physics & Biomedical Engineering, University College London, London, UK
[2] InstaDeep, London, UK
[3] Division of Surgery & Interventional Science, University College London, London, UK
[4] Department of Urology, University College Hospital NHS Foundation Trust, London, UK
[5] Department of Radiology, Stanford School of Medicine, Stanford, California, USA
[6] Department of Urology, Stanford School of Medicine, Stanford, California, USA
[7] Department of Engineering Science, University of Oxford, Oxford, UK
shaheer.saeed.17@ucl.ac.uk

Abstract. The performance of many medical image analysis tasks are strongly associated with image data quality. When developing modern deep learning algorithms, rather than relying on subjective (human-based) image quality assessment (IQA), task amenability potentially provides an objective measure of task-specific image quality. To predict task amenability, an *IQA agent* is trained using reinforcement learning (RL) with a simultaneously optimised *task predictor*, such as a classification or segmentation neural network. In this work, we develop transfer learning or adaptation strategies to increase the adaptability of both the IQA agent and the task predictor so that they are less dependent on high-quality, expert-labelled training data. The proposed transfer learning strategy re-formulates the original RL problem for task amenability in a meta-reinforcement learning (meta-RL) framework. The resulting algorithm facilitates efficient adaptation of the agent to different definitions of image quality, each with its own Markov decision process environment including different images, labels and an adaptable task predictor. Our work demonstrates that the IQA agents pre-trained on non-expert task labels can be adapted to predict task amenability as defined by expert task labels, using only a small set of expert labels. Using 6644 clinical ultrasound images from 249 prostate cancer patients, our results for image classification and segmentation tasks show that the proposed IQA method can be adapted using data with as few as respective 19.7% and 29.6% expert-reviewed consensus labels and still achieve comparable IQA and task performance, which would otherwise require a training dataset with 100% expert labels.

1   Introduction

Medical image quality can influence the downstream clinical tasks intended for medical images [1]. Automated algorithms have been proposed for image quality assessment (IQA), based on human scoring of image quality [2–6], prior clinical knowledge [7, 8] or a set of hand-engineered criteria [9–11]. Task-specific image quality, which measures how well a clinical task can be completed using the image being assessed, may be preferred, but previous methods still rely on human interpretation [2, 3]. When the downstream clinical tasks are completed by automated machine learning algorithms, task-specific IQA may become more relevant, however, human perceived task-specific IQA may not accurately reflect the performance of the machine optimised *task predictors*. Recent works introduce task amenability; defined as the task-specific image quality to directly measure target task performance [12, 13], which also takes into account the dependency between training an automated IQA and the training of a task predictor.

For predicting task amenability for IQA, Saeed *et al.* [12] proposed to train a controller; here, a reinforcement learning (RL) agent, together with the task predictor. Classification and segmentation neural networks were tested as the task predictors. The trained controller predicts significantly different task amenability scores to those determined by humans, with or without requiring human labels of task amenability during training.

By definition, this IQA approach is inevitably dependent on the task predictor and the labelled data used to train such a task predictor, in the case of supervised learning. In clinical practice, the feasibility and cost associated with obtaining quality labelled data sets for various target tasks can not be overlooked. Therefore, we propose a transfer learning strategy to train the IQA agent based on meta-reinforcement learning (meta-RL) across multiple environments. These RL environments can then be designed to reflect different Markov decision processes (MDPs) with differently labelled data. At the same time, a shared task predictor[1] is trained between these MDPs, such that it may be adapted together with the meta-trained controller. Equipping adaptation ability to both the controller and the task predictor has several potential applications for the efficient use of labelled data. In this work, we demonstrate the resulting adaptation ability from relatively low-quality *non-expert* task labels annotated by individual observers to high-quality *expert* labels carefully curated by reviewed consensus.

The contributions of the work are summarised as follows: 1) we propose a transfer learning or adaptation strategy to train an adaptable IQA system; 2) we design a meta-RL algorithm for training the task-amenability-predicting controller together with a target task predictor, which is shared amongst multiple environments, such that training to convergence is not required on every

---

[1] *Tasks* refer to the target classification or segmentation tasks, while MDPs or environments are preferred over *meta-tasks* found in meta-learning literature for clarity.

time-step and where adaptability is equipped to both the inner and outer loops simultaneously; 3) we demonstrate the efficacy of the proposed transfer learning strategy with experiments using a large set of clinical ultrasound images from prostate cancer patients, labelled by four different observers with varying experience and expertise; 4) the experiments show that using 20-30% of the expert labels is sufficient to fine-tune both the RL controller and the task predictor to achieve comparable performances to when they are trained using the full set of expert labels.

2   Methods

2.1   Image quality assessment by task amenable data selection

In this work, we follow the IQA formulation proposed by Saeed *et al.* [12]. There are two parametric functions, a task predictor $f(\cdot; w) : \mathcal{X} \to \mathcal{Y}$ and a controller $h(\cdot; \theta) : \mathcal{X} \to [0, 1]$, with parameters w and θ, respectively. $\mathcal{X}$ and $\mathcal{Y}$ are the respective image and label domains with $\mathcal{P}_{XY}$ being the joint image-label distribution, with a density function $p(x, y)$.

The task predictor f is optimised to predict labels, by minimising the loss function $L_f : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}_{\geq 0}$ using sampled data:

$$\min_{w} \mathbb{E}_{(x,y) \sim \mathcal{P}^h_{XY}} [L_f(f(x; w), y)], \tag{1}$$

where $\mathcal{P}^h_{XY}$ is the controller-selected joint image-label distribution, with density function $p^h(x, y) \propto p(x, y) h(x; \theta)$.

The controller h is optimised to measure image quality (task amenability), by minimising the metric function $L_h : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}_{\geq 0}$:

$$\min_{\theta} \mathbb{E}_{(x,y) \sim \mathcal{P}^h_{XY}} [L_h(f(x; w), y)] \tag{2}$$

where L_h is in general a non-differentiable metric computed on the validation set, and different to L_f.

The optimisation is performed using reinforcement learning, where the environment consists of the training set from $\mathcal{P}_{XY}$ and the task predictor $f(\cdot; w)$; the agent is the controller $h(\cdot; \theta)$ whose action is sample selection $a_t = \{a_{i,t}\}_{i=1}^{B} \in \{0, 1\}^B$, based on the predicted quality scores $\{h(x_i; \theta)\}_{i=1}^{B}$, from a mini-batch of training samples $\mathcal{B}_t = \{(x_i, y_i)\}_{i=1}^{B}$; and the reward is the task predictor performance on a validation set from the same distribution $\mathcal{P}_{XY}$, which is computed after training, for a fixed number of steps, using the selected samples. In this work we use the reward formulation, from [12], which does not require human task amenability labels, and weights the validation set using controller predictions. R_t is thus the reward which is a weighted sum of validation set performance.

2.2   Meta-reinforcement learning with different labels

In this section, we consider multiple label distributions $\{\mathcal{P}^k_{Y|X}\}_{k=1}^{K}$, such that each sample x has multiple labels $\{y^k\}_{k=1}^{K}$. The joint distributions are thereby
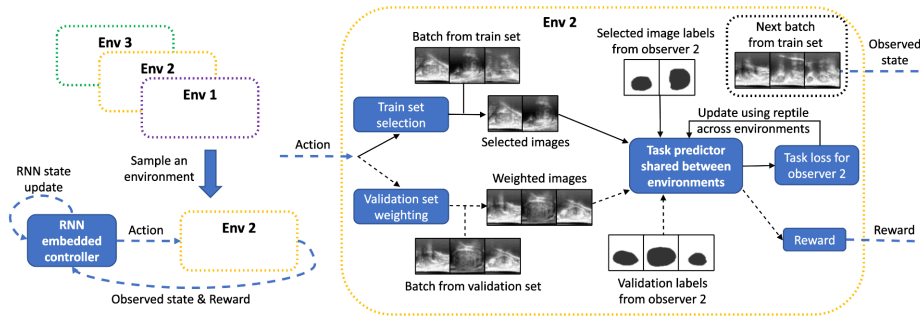
Fig. 1: An overview of the proposed meta-RL framework for training the task predictor and the RNN-embedded controller (the IQA agent).

$\mathcal{P}_{XY}^k = \mathcal{P}_X \mathcal{P}_{Y|X}^k$ for $k = 1, \ldots, K$. Each distribution $\mathcal{P}_{XY}^k$ forms an RL environment with an MDP M_k. These MDPs are assumed to be sampled from the same MDP distribution $\mathcal{P}_M$, i.e. $M_k \sim \mathcal{P}_M$. The task predictors $f(\cdot; w)$ and controller $h(\cdot; \theta)$ are both shared across different environments.

We adopt the meta-RL formulation [14, 15] for reinforcement learning across multiple environments. Given a set of MDPs $\{M_k\}_{k=1}^K$, a *trial* is defined as multiple episodes with a sampled MDP M_k. The meta-RL agent learns across multiple trials by sampling $M_k \sim \mathcal{P}_M$. Different from the RL with one single fixed environment, at time $t + 1$, the meta-RL agent h takes the action a_t, raw reward r_t, and termination flag d_t at the previous time step in addition to the observed current state s_{t+1}. Note that for per-sample operation $r_t = R_t$ at the episode end, and zero otherwise, similar to sparse reward formulations in [14, 15]. Denote the input tuple as $\tau_{t+1} = (s_{t+1}, a_t, r_t, d_t)$, thereby $h(\cdot; \theta)$ is now defined with a space of $\mathcal{X} \times [0, 1] \times \mathbb{R} \times \{0, 1\}$.

In this work, the meta-RL agent adopts a recurrent neural network (RNN) with internal memory shared across episodes in the same trial. Importantly, the internal memory is reset when a trial finishes, i.e. before another environment is sampled. This mechanism allows test-time adaptability, even with fixed weights [16–20], and thereby transfers knowledge between environments [14, 15, 21]. This is due to the RNN making the controller a function of the history leading up to a sample such that changing history can influence the action for that sample. The full algorithm is described in Algorithm 1, with details for configuring episodic mini-batches and meta-loop trials. An overview is also presented in Fig. 1. In our implementation, proximal policy optimisation (PPO) [22] was used to train the controller. The task predictor employs the Reptile scheme [23] to allow potential data efficiency benefit for adapting to different observer labels. The predictor is updated in two steps: 1) update starting weights w_{t+1} of predictor $f(\cdot; w_{t+1})$ to $w_{t+1,\text{new}}$, using gradient descent based on $\mathcal{B}_{t,\text{selected}}$; 2) update weights using $w_{t+1} \leftarrow w_{t+1} + \epsilon(w_{t+1,\text{new}} - w_{t+1})$ where ϵ is 1.0 initially and is linearly annealed to 0.0 as trial iterates. It is worth noting that the IQA algorithm from [12] can be considered a special case of our proposed method with only one environment.

After training using the scheme described in Algorithm 1, the adaptation stage, for both the controller and task-predictor, can be performed on a single MDP of interest $M_a \sim \mathcal{P}_M$, where M_a is the environment which we would like to adapt to. If multiple iterations of the outer loop are required, the internal state of the controller is only reset on the first iteration. The controller weights remain fixed; adaptability is a result of updating internal state.

Algorithm 1: Adaptable image quality assessment by task amenability

---

Data: Multiple MDPs $M_k \sim \mathcal{P}_M$.
Result: Task predictor $f(\cdot; w)$ and controller $h(\cdot; \theta)$.

while *not converged* **do**
    Sample an MDP $M_k \sim \mathcal{P}_M$;
    Reset the internal state of controller h;
    for *Each episode in all episodes* **do**
        for $t \leftarrow 1$ **to** T **do**
            Sample a training mini-batch $\mathcal{B}_t = \{(x_{i,t}, y_{i,t})\}_{i=1}^{B}$;
            Compute selection probabilities $\{h_{i,t}\}_{i=1}^{B} = \{h(\tau_{i,t}; \theta_t)\}_{i=1}^{B}$;
            Sample actions $a_t = \{a_{i,t}\}_{i=1}^{B}$ w.r.t. $a_{i,t} \sim \text{Bernoulli}(h_{i,t})$;
            Select samples $\mathcal{B}_{t,\text{selected}}$ from $\mathcal{B}_t$;
            Update predictor $f(\cdot; w_t)$ with $\mathcal{B}_{t,\text{selected}}$ using Reptile;
            Compute reward R_t;
        end
        Collect one episode $\{\mathcal{B}_t, a_t, R_t\}_{t=1}^{T}$;
        Update controller $h(\cdot; \theta)$ using the RL algorithm PPO;
    end
end

---

3 Experiments

In this work we use 6644 2D ultrasound images from 249 prostate cancer patients. During the early stages of ultrasound-guided biopsy procedures, images were acquired using a transperineal ultrasound probe (C41L47RP, HI-VISION Preirus, Hitachi Medical Systems Europe) as part of SmartTarget: THERAPY and SmartTarget: BIOPSY clinical trials (clinicaltrials.gov identifiers NCT02290561 and NCT02341677 respectively). Images from each subject initially consisted of 50-120 frames. For feasibility of manual labelling, frames were sampled at four-degree intervals where relative rotation angles were tracked using a digital transperineal stepper (D&K Technologies GmbH, Barum, Germany). The resulting 6644 2D ultrasound images were randomly split, at the patient level, into training, validation and holdout sets, with 4429, 1092 and 1123 images from 174, 37 and 38 subjects, respectively.

Three sets of task label $\{L_i\}_{i=1}^3$ were collected from three trained biomedical engineering researchers. These individually-labelled are referred to as "non-expert" label sets for brevity. In addition, the fourth set of "expert" labels L_* was curated by a urologist, first carefully reviewing a reference set of consensus labels and then editing them as deemed necessary. For all label sets, each image has both a binary label indicating prostate presence for classification and a binary mask of the prostate gland for segmentation.

The task predictor algorithms used for the two tested applications are the same as [12]. For classification, AlexNet [24] was used with a cross-entropy loss and a reward based on classification accuracy. For segmentation, U-Net [25] was used with a pixel-wise cross-entropy loss and a reward based on mean binary Dice score. The controller had a three-layer convolutional encoder, before feeding the encoded features to an RNN with a stacked-LSTM architecture, as described in [15]. Experimental results are reported for empirically configured networks and default hyper-parameter values remain unchanged unless specified.

The following three different IQA models were trained and compared.

- *Baseline*: Trained with all training and validation data using only the high-quality expert labels L_*. That is, only one "expert-labelled" environment in training, establishing a reference for achievable IQA system performance.
- *Meta-RL*: The proposed model that was first trained with training and validation data using the non-expert labels $\{L_i\}_{i=1}^3$ as three different environments. Both the task predictor and the controller were subsequently adapted with $k \times 100\%$ training and validation data using the expert labels L_*.
- *Meta-RL Variant*: For comparison, a basic implementation of transfer learning. The model was first trained with all training and validation data using the shuffled non-expert labels $\{L_i\}_{i=1}^3$ as one single environment, i.e. without considering different environment-specific trials, and the Reptile update for optimising the task predictor reduced to standard gradient descent. Adaptation was done with $k \times 100\%$ training and validation data using the expert labels L_*. The internal state of RNN was not reset before fine-tuning.

We evaluate the IQA models jointly with the task predictors using task performance, which serves as both a direct evaluation of the task-predictor and an indirect evaluation of the IQA agent by its task amenability definition. We report mean accuracy (Acc.) and mean binary Dice score (Dice) on the holdout set using expert labels for classification and segmentation, respectively. These measures are averaged over all 2D slices in the holdout set. Where controller selection is used, the metric is computed over the selected samples only. Samples are selected by rejecting the subset with the lowest controller predicted values, with the specified rejection ratios. Standard deviation (St.D) is reported to measure inter-patient variance, with which, paired t-test results with a significance level of 5% are reported when any comparison is made. We evaluate the models for varying k-values, where k is the ratio of expert-labelled samples used for adaptation ($k \times 100\%$ samples used).

4 Results

Table 1: Comparison of holdout set results with a rejection ratio set to 5%

Tasks		Prostate Classification (Acc.)	Prostate Segmentation (Dice)
IQA Methods	k	Mean ± St.D.	Mean ± St.D.
Baseline	N/A	0.932 ± 0.011	0.894 ± 0.016
Meta-RL	0.5	0.936 ± 0.012	0.892 ± 0.018
	0.4	0.929 ± 0.016	0.886 ± 0.014
	0.3	0.926 ± 0.010	0.888 ± 0.020
	0.2	0.925 ± 0.017	0.873 ± 0.017
	0.1	0.911 ± 0.012	0.863 ± 0.020
	0.0	0.908 ± 0.010	0.857 ± 0.018
Meta-RL Variant	0.5	0.931 ± 0.015	0.884 ± 0.016
	0.4	0.920 ± 0.010	0.882 ± 0.021
	0.3	0.919 ± 0.013	0.882 ± 0.015
	0.2	0.916 ± 0.014	0.860 ± 0.014
	0.1	0.905 ± 0.014	0.858 ± 0.021
	0.0	0.896 ± 0.016	0.849 ± 0.017

The proposed meta-training took, on average, approximately 48 hours and the meta-testing (model fine-tuning) took 1-2 hours on a single Nvidia Quadro P5000 GPU. This result reflects the design of the proposed adaptation strategy for data efficiency and, arguably, also for computational efficiency.

Performance of the IQA models, in terms of Acc. and Dice, are summarised in Table 1 and plotted in Fig. 2 against varying k values. In the prostate presence classification task, no statistical significance was found between the baseline and meta-RL for k values from 0.5 to 0.2 (p-values ranged from 0.10 to 0.23). However, meta-RL performance for low k values, $k = 0.1$ or 0.0, was significantly lower than that of the baseline ($p < 0.01$ for both). In the prostate segmentation task, no statistical significance was found between the two, for k-values from 0.5 to 0.3 (p-values ranged from 0.07 to 0.17), but a significantly lower performance was found for meta-RL for low k values from 0.2 to 0.0 ($p<0.01$ for all).

For the ablation study comparing meta-RL to the meta-RL variant, the proposed meta-RL framework generally outperformed the meta-RL variant for the same k values, for both tested target tasks, as detailed in Table 1. For classification, we report a statistically significant difference between the two, for the same k values from 0.0 to 0.4 ($p<0.01$ for all), while no significance was found when the k increased to 0.5 ($p=0.06$). For segmentation, superior performance from the proposed meta-RL was statistically significant for all k values ($p<0.03$ for all). From an ablation study, with and without the Reptile scheme for updating task predictors, the Reptile-omitted meta-RL classification and segmentation tasks achieved Acc.$=0.901 \pm 0.013$ and Dice$=0.851 \pm 0.013$, respectively, when $k = 0$. The improvement, when using the Reptile scheme, was statistically significant with $p<0.01$ for both, but no significant difference was found for other k values.

Fig. 3 provides visual examples of selection decisions by the adapted IQA agent. With 5% rejection ratio, all these rejected examples seem visually challenging for respective classification and segmentation tasks, and rejecting these examples improved performances of the simultaneously learned task predictors.
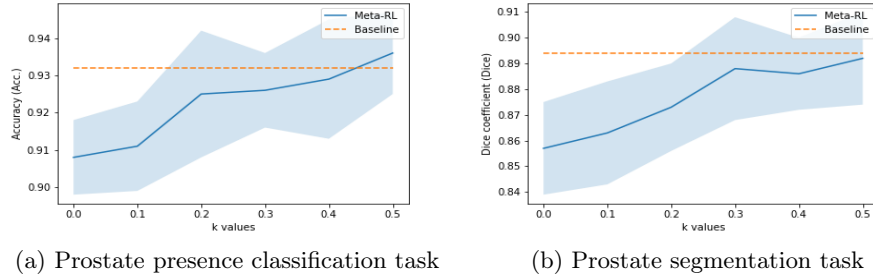

(a) Prostate presence classification task   (b) Prostate segmentation task

Fig. 2: Task performance (in respective Acc. and Dice metrics) against the k values with rejection ratio set to 5%.

(a) Prostate classification task   (b) Prostate segmentation task

Fig. 3: Examples of controller selected and rejected images (rejection ratio=5%) for both tasks. **Blue:** rejected samples; **Red:** selected samples; **Yellow:** rejected samples despite no apparent artefacts or severe noise; **Green:** selected samples despite present artefacts or low contrast. **Orange arrows:** visible artefacts; **Cyan arrows:** regions where gland boundary delineation may be challenging.

5   Discussion and Conclusion

Based on results reported in Sect. 4, for the tested ultrasound guidance application, the proposed adaptation strategy allows for the IQA agent and task predictor to be adapted using as few as 1087 and 1634 expert-labelled images from 42 and 63 subjects (training and validation sets), for classification and segmentation, respectively. Compared with a total of 5521 expert-labelled images from 211 subjects that were required to train the baseline, this is a substantial reduction, to 20-30%, in the required quantity of high-quality and often expensive expert-labelled data. The proposed model also used non-expert labels for training but these may be used for different IQA definitions, further economic analysis is beyond the scope of this work. An adaptable IQA algorithm has been presented, which can be efficiently adapted with new labelled data. The proposed algorithm may have general applicability to alleviate demand for large quantities of training data, for example, for other imaging protocols or target tasks.

References

[1]   L.S. Chow and R. Paramesran. "Review of medical image quality assessment". In: *Biomed. Signal Processing and Control* 27 (2016), pp. 145–154.
[2]   S.J. Esses, X. Lu, T. Zhao, K. Shanbhogue, B. Dane, M. Bruno, and H. Chandarana. "Automated image quality evaluation of T2-weighted liver MRI utilizing deep learning architecture". In: *Journal of Magnetic Resonance Imaging* 47.3 (2018), pp. 723–728.
[3]   G.T. Zago, R.V. Andreão, B. Dorizzi, E. Ottoni, and T. Salles. "Retinal image quality assessment using deep learning". In: *Computers in Biology and Medicine* 103 (2018), pp. 64–70.

[4]   Z.M.C. Baum, E. Bonmati, L. Cristoni, A. Walden, F. Prados, B. Kanber, D.C. Barratt, D.J. Hawkes, G.J.M. Parker, C.A.M.G. Wheeler-Kingshott, and Y. Hu. "Image quality assessment for closed-loop computer-assisted lung ultrasound". In: *Medical Imaging 2021: Image-Guided Procedures, Robotic Interventions, and Modeling*. Ed. by Cristian A. Linte and Jeffrey H. Siewerdsen. Vol. 11598. International Society for Optics and Photonics. SPIE, 2021, pp. 160–166. DOI: `10.1117/12.2581865`.

[5]   A. H. Abdi, C. Luong, T. Tsang, G. Allan, S. Nouranian, J. Jue, D. Hawley, S. Fleming, K. Gin, J. Swift, R. Rohling, and P. Abolmaesumi. "Automatic Quality Assessment of Echocardiograms Using Convolutional Neural Networks: Feasibility on the Apical Four-Chamber View". In: *IEEE Transactions on Medical Imaging* 36.6 (2017), pp. 1221–1230. DOI: `10.1109/TMI.2017.2690836`.

[6]   Z. Liao, H. Girgis, A. Abdi, H. Vaseli, J. Hetherington, R. Rohling, K. Gin, T. Tsang, and P. Abolmaesumi. "On modelling label uncertainty in deep neural networks: automatic estimation of intra-observer variability in 2d echocardiography quality assessment". In: *IEEE Transactions on Medical Imaging* 39.6 (2019), pp. 1868–1883.

[7]   L. Wu, J. Cheng, S. Li, B. Lei, T. Wang, and D. Ni. "FUIQA: Fetal Ultrasound Image Quality Assessment With Deep Convolutional Networks". In: *IEEE Trans. on Cybernetics* 47.5 (2017), pp. 1336–1349.

[8]   Z. Lin, S. Li, D. Ni, Y. Liao, H. Wen, J. Du, S. Chen, T. Wang, and B. Lei. "Multi-task learning for quality assessment of fetal head ultrasound images". In: *Medical Image Analysis* 58 (2019), p. 101548. ISSN: 1361-8415. DOI: `https://doi.org/10.1016/j.media.2019.101548`.

[9]   H. Davis, S. Russell, E. Barriga, M. Abramoff, and P. Soliz. "Vision-based, real-time retinal image quality assessment". In: (2009), pp. 1–6.

[10] T. Köhler, A. Budai, M. F. Kraus, J. Odstrčilik, G. Michelson, and J. Hornegger. "Automatic no-reference quality assessment for retinal fundus images using vessel segmentation". In: *Proc. of the 26th IEEE Int. Symp. on Computer-Based Medical Systems*. 2013, pp. 95–100.

[11] C.P. Loizou, C.S. Pattichis, M. Pantziaris, T. Tyllis, and A. Nicolaides. "Quality evaluation of ultrasound imaging in the carotid artery based on normalization and speckle reduction filtering". In: *Med. and Bio. Eng. and Comp.* 44 (2006).

[12] Shaheer U. Saeed, Yunguan Fu, Zachary M. C. Baum, Qianye Yang, Mirabela Rusu, Richard E. Fan, Geoffrey A. Sonn, Dean C. Barratt, and Yipeng Hu. "Learning Image Quality Assessment by Reinforcing Task Amenable Data Selection". In: *Information Processing in Medical Imaging*. Ed. by Aasa Feragen, Stefan Sommer, Julia Schnabel, and Mads Nielsen. Cham: Springer International Publishing, 2021, pp. 755–766.

[13] J. Yoon, S. Arik, and T. Pfister. *Data Valuation using Reinforcement Learning*. 2020. arXiv: `1909.11671`.

[14]   Y. Duan, J. Schulman, X. Chen, P.L. Bartlett, I. Sutskever, and P. Abbeel. *RL²: Fast Reinforcement Learning via Slow Reinforcement Learning*. 2016. arXiv: `1611.02779 [cs.AI]`.

[15]   J.X. Wang, Z. Kurth-Nelson, D. Tirumala, H. Soyer, J.Z. Leibo, R. Munos, C. Blundell, D. Kumaran, and M. Botvinick. *Learning to reinforcement learn*. 2017. arXiv: `1611.05763 [cs.LG]`.

[16]   N.E. Cotter and P.R. Conwell. "Fixed-weight networks can learn". In: *1990 IJCNN International Joint Conference on Neural Networks* (1990), 553–559 vol.3.

[17]   A. Santoro, S. Bartunov, M. Botvinick, D. Wierstra, and T. Lillicrap. "Meta-Learning with Memory-Augmented Neural Networks". In: *Proceedings of The 33rd International Conference on Machine Learning*. Ed. by M.F. Balcan and K.Q. Weinberger. Vol. 48. Proceedings of Machine Learning Research. New York, New York, USA: PMLR, 2016, pp. 1842–1850.

[18]   A.S. Younger, P.R. Conwell, and N.E. Cotter. "Fixed-weight on-line learning". In: *IEEE Transactions on Neural Networks* 10.2 (1999), pp. 272–283. DOI: `10.1109/72.750553`.

[19]   S. Hochreiter, A.S. Younger, and P.R. Conwell. "Learning to learn using gradient descent". In: *International Conference on Artificial Neural Networks*. Springer. 2001, pp. 87–94.

[20]   D.V. Prokhorov, L.A. Feldkarnp, and I.Y. Tyukin. "Adaptive behavior with fixed weights in RNN: an overview". In: *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02*. Vol. 3. 2002, 2018–2022 vol.3. DOI: `10.1109/IJCNN.2002.1007449`.

[21]   M. Botvinick, S. Ritter, J.X. Wang, Z. Kurth-Nelson, C. Blundell, and D. Hassabis. "Reinforcement learning, fast and slow". In: *Trends in cognitive sciences* 23.5 (2019), pp. 408–422.

[22]   J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov. *Proximal Policy Optimization Algorithms*. 2017. arXiv: `1707.06347 [cs.LG]`.

[23]   A. Nichol, J. Achiam, and J. Schulman. *On First-Order Meta-Learning Algorithms*. 2018. arXiv: `1803.02999 [cs.LG]`.

[24]   A. Krizhevsky, I. Sutskever, and G. Hinton. "Imagenet classification with deep convolutional neural networks". In: *NeurIPS* (2012).

[25]   O. Ronneberger, P. Fischer, and T. Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *MICCAI*. Vol. 9351. Springer. 2015.