# Deep Reinforcement Learning in Medical Imaging: A Literature Review

S. Kevin Zhoua, Hoang Ngan Leb, Khoa Luub , Hien V. Nguyenc , Nicholas Ayached

JOCELYN BADURIA

# Table of Contents

# Abstract


- What is Deep Reinforcement Learning(DRL)?

  - It is a subset of Machine Learning called Deep Learning with added top layered Implementation process called Reinforcement learning.

- What is Reinforcement Learning(RL)?

  - It is a learning process that uses a game theory analogy, that it learns through a series of actions through level up reward system.

- Most recent applications of DRL have shown tremendous impact on healthcare and medicine field. DRL is very useful in learning medical images analysis for "object/lesion detection, surgical image segmentation/augmentation, registration of significant medical images, view plane image localization, personal mobile health intervention".

# Introduction

- DRL is a (SOTA) State-of-the-Art framework in Artificial Intelligence.
  - It scales up the learning sequential decision process and can highly learn with delayed supervised information.
  - It can also deal with maximizing the "non-differentiable" metrics of parametric setting of classification accuracy problem.
  - It also speed up the searching detection process for optimal transformation of image registration accuracy.
  - It improves the memory allocation of high-dimensional images. Small images can be used in DRL object detection process.
- This survey paper expanded its study on many (SOTA) state-of-the-art contents.

# Basics of Reinforcement Learning

- Markov Decision Process (MDP)

  - Is the essential foundation theory of Reinforcement Learning, a process "whose the conditional probability distribution of future states only depends upon the present state".

  - The introduction of reward system not only depends on the current state but also, on the action that leads to future state.

  - One good example of this is the famous online game "Fortnite". Usually, the game player chooses an action while playing according to the set of rules and the environment receives the action and produces a reward. Then the player transfers to the next level until he/she reached the final level of the game which is the future state.

# Basics of Reinforcement Learning

- MDP is defined into five elements:
    1. set of state/observation space
    2. set of actions the agents can choose from
    3. transition probability function
    4. reward function
    5. discount factor
- The main objective of (RL) Reinforcement Learning is to find the prime rules or best- known solution that gives the highest “expected cumulative reward called return - the weighted sum of immediate rewards”.

# Basics of Reinforcement Learning

- Reinforcement Learning is divided into two functions

 1. Model-Free Methods

 - This method is the learning process of a gamer during the interaction process with other players and the environment to exceed the maximum policy.

 2. Model-Based Methods

 - This method is when the experienced gamer knows such a certain rules or game policy and use it for its own advantage to exceed the maximum policy.

# Introduction to Deep Reinforcement Learning

- Autoencoder

  - Is a representation unsupervised learning used for “dimensionality reduction and feature selection”.

- Deep Belief Network and Deep Autoencoders

  - Are unsupervised learning used for network initialization.

  - Deep autoencoders have one seen input layer and one invisible layer.

  - Deep belief network is based on “Restricted Boltzmann Machines” which contains of layered input data and a layered invisible units that can gain a knowledge  to describe the features that are gathered from the “higher-order of correlations in the data”.

# Introduction to Deep Reinforcement Learning

- Multi-layer Perceptron (MLP)

  - Neuron is the simplest form of neural network.

  - Perceptron is a computational single form of neuron that has one or many inputs, main-processor, and one output.

  - There are two types of Neural Networks.

    1. Convolutional Neural Networks (CNN)
    2. Recurrent Neural Networks (RNN)

- Convolutional Neural Networks (CNN)

  - A fully joined multi-layer perceptrons that perform weight sharing during the data processing in a grid like topology network.

# Introduction to Deep Reinforcement Learning

- Recurrent Neural Network (RNN)

  - A sequential model that performs the same task for each sequence on which the output depends on the previous task computations.

  - Long Short-Term Memory (LSTM) was born to address the issue of the difficulty of (RNN) training for long-term dependencies. It was abled to maintain the memory that updates and release its content when it is needed during training process.

  - Gated Recurrent Unit (GRU) was proposed recently to improve the process of capturing dependencies at different time intervals. The difference with LSTM it has no separate memory cells.

# Introduction to Deep Reinforcement Learning

- Model-free DRL Algorithms

  - There are three approaches for this algorithms namely

    1. Value-Based DRL Methods
    2. Policy Gradient DRL Methods
    3. Actor-Critic DRL Methods

- Value-Based DRL Methods

  1. Deep Q-Learning Network (DQN)

     - Is the popular one among the three. This method directly knows the policy knowledge from the higher dimension inputs in a deep neural network.

     - This uses a regression modelling but with main limitation that Q* values tends to over-estimate because of max values in MSE loss computation.

$$y(s_t, a_t) = R(s_t, s_{t+1}) + \gamma \max_{a_{t+1}} Q^*(s_{t_1}, a_{t+1}, \theta_t)$$

# Introduction to Deep Reinforcement Learning

- Value-Based DRL Methods

2. Double DQN is the improved DQN version that uses double implementation of DQN. It addresses the over-estimation issue with its improved estimation but very costly process although the easiest one method.

$$y_1 = R(s_t, s_{t+1}) + \gamma Q_1^*(s_{t+1}, \arg\max_{a_{t+1}} Q_2^*(s_{t+1}, a_{t+1}; \theta_2); \theta_1),$$
$$y_2 = R(s_t, s_{t+1}) + \gamma Q_2^*(s_{t+1}, \arg\max_{a_{t+1}} Q_1^*(s_{t+1}, a_{t+1}; \theta_1); \theta_2).$$

3. Dueling DQN addresses the issue of DQN on its value *V**, instead of lowering the value it remembers only the low reward by updating the Q* value. It introduces the advantage function.

$$A^\pi(s_t, a_t) = Q^\pi(s_t, a_t) - V^\pi(s_t).$$

Therefore, Q-value is rewritten as

$$Q^*(s, a) = A^*(s, a) + V^*(s)$$

$$Q^*(s_t, a_t) = V^*(s_t) + A^*(s_t, a_t) - mean_{a_{t+1}} A^*(s_t, a_{t+1}).$$

# Introduction to Deep Reinforcement Learning

- Value-Based DRL Methods

4. Deep Recurrent Q-Network (DRQN)

- This addresses the limitation of memory and imperfect information for each decision point. It added the Recurrent Neural Network (RNN) into DQN by replacing its first DQN fully joined layer.

- This is the most famous improvement of DQN by adding "one-step approximation with N-steps".

# Introduction to Deep Reinforcement Learning

- Policy Gradient DRL Methods

  - It differs from the Value-based DRL methods. It mainly optimizes the gradient objective function.

$$\mathcal{G}(\theta) = \mathbb{E}_{\mathcal{T}\sim\pi_\theta} \sum_{t=1} \gamma^{t-1} R(s_{t-1}, s_t) \to \max_\theta .$$

  - REINFORCE was introduce by using Monte-Carlo estimation to approximately calculate gradient.

  - This method can be applied into any parametric problems however it is time costly for convergence for local optima.

  - The naive REINFORCE “variance gradient estimation” was address by adding the “variance reducing technique” not to affect the outcome of expected result.

# Introduction to Deep Reinforcement Learning

- Actor-critic DRL Algorithm

  - In comparison with the two methods, value-based and policy gradient, policy gradient methods out-performed the value based for its fast and continuous training convergence. But the value-based method are steadier and exhibits sample efficiency.

  - These architectures are fully optimized by Actor-critic algorithm.

  - The main idea of actor-critics is to divide the model into two parts namely:

    1. Computing an action based on a state .
    2. Producing the Q value of the action.

# Introduction to Deep Reinforcement Learning

- Actor-critic DRL Algorithm

  - Best example of this is a kid/mother relationship. A kid explored the space environment, and the mother watches her kid. If the kid does not behave in an environment, the mother tells the kid to behave appropriately. Until such time kids knows the rule and understand what is the difference of good and bad behavior.

  1. Advantage actor-critic (A2C)

  - Consists of two neural network, an actor network - policy and a critic network.

  2. Asynchronous advantage actor critic(A3C) is another strategy to implement the actor critic agent to meet the efficiency of memory.

# Introduction to Deep Reinforcement Learning

- Model Based Algorithms

  - Is an experience learned model from a function approximation. Theoretically this model does not require based knowledge, but it may help if the based knowledge is added which may help the training process converge faster.

  - Most common model-based approach includes value function and policy search methods.

  1. Value Function

     - Used Convolutional Neural Network (CNN) to approximate Q* value function in high dimensional space.

  2. Policy Search

     - It directly aims to find the policy by means of gradient (free or based) methods.

# Introduction to Deep Reinforcement Learning

- Useful techniques to train an agent

1. Experience Replay

   - Based on the experience of a gamer and a useful part in rules policy learning.

2. Minibatch Learning

   - With experience replay and minibatch learning it allows the training in more batches thus it helps the learning more robust from noise.

# Introduction to Deep Reinforcement Learning

- Useful techniques to train an agent

  3. Target Q-network Freezing

     - Two networks are used; the one that interacts and generate target Q* values to compute losses and the other plays as a target network and its weights are fixed and updated by the first network.

  4. Reward clipping

     - Clipping the reward to capture the learned process in a scalable way.

# DRL in Medical Imaging

- DRL is mostly used in medical imaging analysis.

  - Application varies from these field of analysis :

    1. Landmark Detection
    2. Image Registration, Segmentation
    3. Object Lesion localization and classification
    4. View Plane localization
    5. Plaque Tracking
    6. Vessel Extraction

  - It is also used in hyperparameter tuning, data augmentation and neural search architecture.

  - Mostly shared non-differentiable optimization for solution.

# DRL in Medical Imaging

- DRL for parametric Medical Image Analysis

  - This model estimated a low dimensional image for analysis.

  - There are three elements that is required to formulate a DRL framework.

  1. Action

     - The agent takes the image by moving the ith parameter independently in certain value while keeping the other parameters in same position.

  2. State

     - The state defined in environment is image at certain points – centered position.

  3. Reward

     - When the target signal is hit or closer the reward function provides a stimuli signal to the agent.

# DRL in Medical Imaging

- DRL for parametric Medical Image Analysis

  - Once the three elements are available the DQL algorithm was invoked and triggered to learn the process.

  - Then the process learned, the Q-function loss is calculated using the maximum value.

  - Using the greedy search fashion for path supervision it maximizes the reward for every iteration process.

  - This eventually leads to convert the Reinforcement Learning (RL) into Supervised Learning model. From there we can start learning a model using classification or regression function.

# DRL in Medical Imaging

- Landmark Detection

  - This landmark anatomical structures plays an important roles in navigating the image sample. It is just like a travelling geographical landmarks in map.

  - It helps the tourist to explore the places they wanted to visit.


**Fig. 9. The list of 49 anatomical landmarks. Courtesy of Ghesu et al. (2018).**

1. Artificial Agent – presents the multi-scale approach for detection of anatomical landmarks in 3Dimensional space. It exhibits perfect detection with no false positive and negative values.

# DRL in Medical Imaging

- Landmark Detection

2. Supervised action classifier – image partitioning of landmark detection. This is taken from the path supervision approach.

- This modeling allow four possible action types that helps to achieve the best result accuracy of image detection from a "cardiac arrest or obstetric ultrasound image in two datasets with 1323 and 1642 patients with compared to SOTA - Artificial Agent.

$$
\begin{aligned}
UP: &\quad (d_x^{(0)} = 0,\ d_y^{(0)} = -1),\\
RIGHT: &\quad (d_x^{(1)} = 1,\ d_y^{(1)} = 0),\\
DOWN: &\quad (d_x^{(2)} = 0,\ d_y^{(2)} = 1),\\
LEFT: &\quad (d_x^{(3)} = -1,\ d_y^{(3)} = 0).
\end{aligned}
$$

Pixel-wise optimal action step with slopes +/-1 crossing the landmark

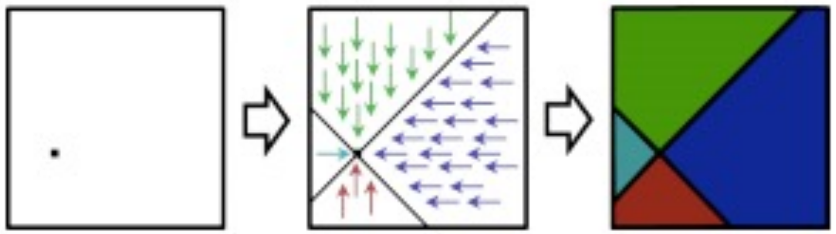

**Fig. 11. The discrete action map representation.**

# DRL in Medical Imaging

- Image Registration

  - For comparison in different times and modalities, robust image registration in medical imaging is needed.

  There are two ways used for this modelling.

  1. Rigid Registration – It uses a Deep Convolutional Neural Network (DCNN) that utilizes the path supervision method to end-to-end training.

  - This modelling is evaluated into two datasets (spine – 87 pairs of images, heart - 97 pairs of images). This artificial agent outperforms SOTA(State of the Art) methods with big margin in terms of accuracy and robustness.

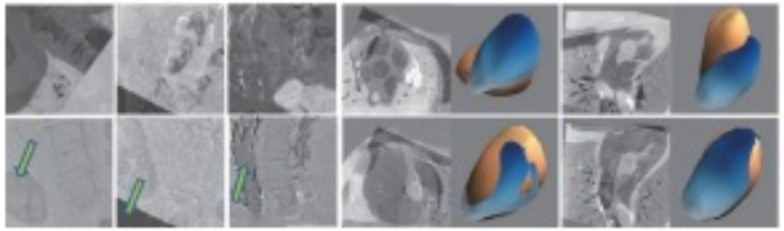

Fig. 12. Registration examples shown as the difference between the reference and floating images, before (upper row) and after (lower row) registration. The mesh overlay before and after registration is shown for cardiac use case for improved visualization. Picture courtesy of Liao et al. (2017).

# DRL in Medical Imaging

- Image Registration

1. Rigid Registration

- Updated model using the dueling DQN by learning the Q function instead of path supervision was introduced to previous method achieves the SOTA when compared with another model.

2. Non-rigid Registration

- A non-rigid registration comes into full picture when the rigid registration performs an insufficient transformation between two images.

# DRL in Medical Imaging

- Image Registration

2. Non-rigid Registration

- Studying the parametric Statistical deformed pair images for an "organ-centered" MR registered prostate images resulted to better performance when compared to other SOTA (State of the Art) models. The median DICE coefficients is 0.96 for 2D and 0.88 in 3D. The Q function thus learned from this study.

- This was also concluded when tested with image registration prostate MR data (41 3D images, 8 testing samples resulting to 56 inter-subject image pairs). Both results are better compared with other state of the art registration model.

# DRL in Medical Imaging

- Object/lesion Localization and Detection
  - Studies were also concluded for this model for detecting breast lesions taken from (DCE-MRI) Dynamic Contrast-Enhanced Magnetic Resonance Imaging machine.
  - Bounding box was defined with its actions together with signal reward function
  - Deep Q-Learning Network (DQN) is used based on ResNet Architecture.
  - Samples were taken from 117 patients with training set 58 patients (annotated 72 lesions ), testing set from 59 patients (69 annotated lesions).
  - Accuracy shows similar results with state-of-the-art model but with biggest impact on speed with time reduced detection.


**Fig. 13. The illustration of the detection process, with the learnt DRL agent outputting a series of allowable actions to realize final detection of a 3D lesion. Picture courtesy of Maicas et al. (2017).**

# DRL in Medical Imaging

- Object/lesion Localization and Detection
  - Other studies also took the challenges of big computational histopathological images of breast cancer classification.
  - Pathologist usually selects an abnormal area of the breast and study its details.
  - Such mechanisms captures the attention to improve the selection and classification of abnormal breast region.
  - DRL was used for two task, one for selection and the other one for classification. Selection is training the certain area of image. The image selected is used for classification.
  - This model achieves its 98% accuracy while consuming only 50% of the old Pathological method, the attention-based approach.

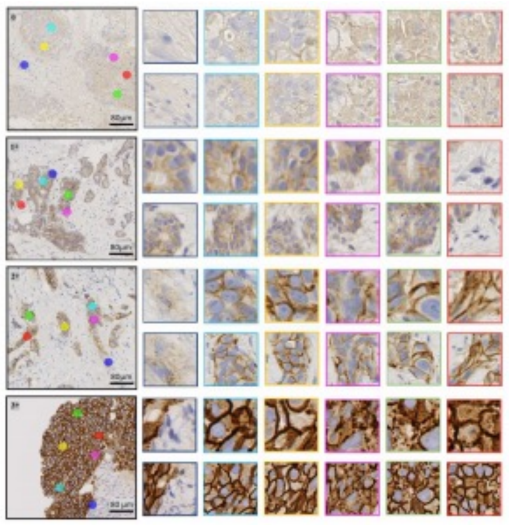

Fig. 14. Example of four image tiles with selected regions-of-interest (ROIs) predicted by Bae et al. (2019), for each HER2 score (0-3+), respectively. The first column shows the input images and colored disks shows the predicted locations. The remaining columns show the selected regions at 40× and 20× around the selected locations. The first selected region is shown with blue bounding boxes and the last selected region is shown with red bounding boxes. Picture courtesy of Bae et al. (2019).

# DRL in Medical Imaging

- View Plane Localization

  - New method proposed for detection of canonical view in cardiac and MR brain images.
  - Studies used the DQN, DDQN, Duel DQN and Duel DDQN.
  - A 3D plane is parameterized in 4D vector, with signed reward function and action steps coarse-to-fine selection.
  - Similarly other method added the image augmentation starting warm for better initialization with active module termination.
  - This added approach improves the efficiency and accuracy of plane localization based on in-house datasets of 430 prenatal 3D ultrasound volume of fetal head.

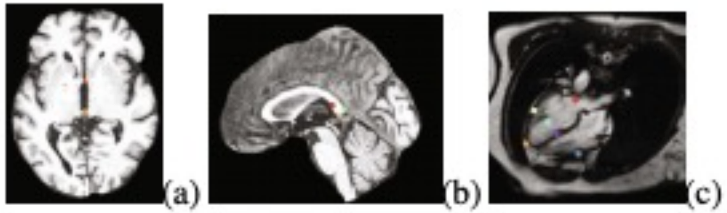


Fig. 15. The viewing planes for detection: (a) Brain axial ACPC plane, (b) Brain mid saggital plane (c) Cardiac apical four chamber plane. The landmarks are visualized for better definition of the plane and used for error calculation. Picture courtesy of Alansary et al. (2018).

# DRL in Medical Imaging

- Plaque Tracking
  - Plaque Atherosclerosis monitoring is being explored by DRL.
  - Plaque is composed of cholesterol, fat, calcium, and other substances found in the blood. This buildup can lead to serious health problems or even death.
  - Still this proposal is having some challenges due to various intravascular morphology.

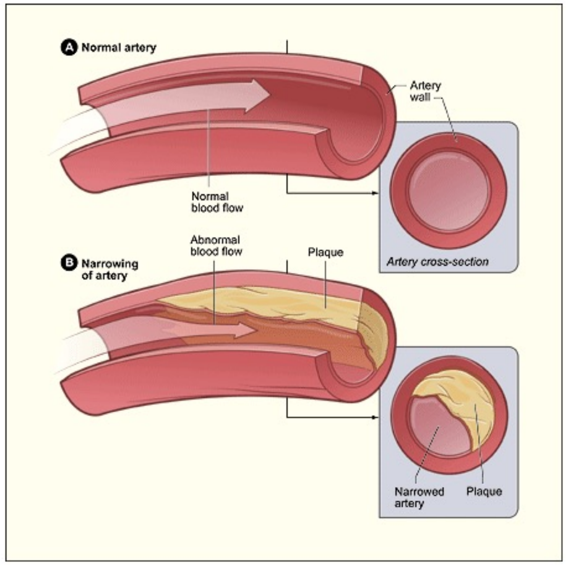


Fig A. Plaque

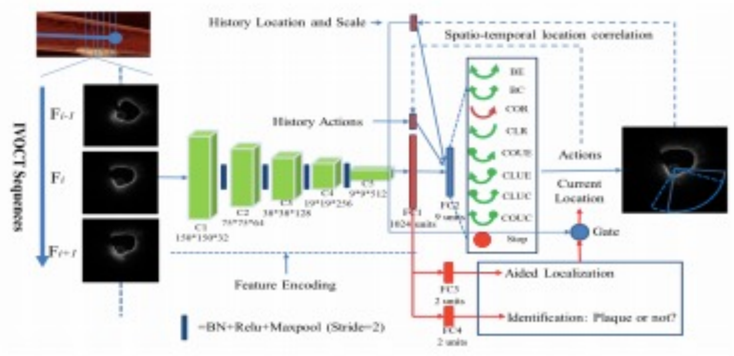


**Fig. 16. The DRL framework is proposed to leverage the spatiotemporal information to achieve continuous and accurate plaque tracking. Picture courtesy of Luo et al. (2019).**

# DRL in Medical Imaging

- Vessel Centerline Extraction

  - DRL using navigation model tracing the vessel centerline.

    A point to curve measure is defined in two terms

    - Pulls agent position towards the true center label
    - The agent is force towards the direction of curve.
    - This method achieves good performance compared with 3D CNN supervised learning.
    - The method was updated using DDQN with 3D CNN to improve the centerline accuracy with speed of 7seconds of inference.

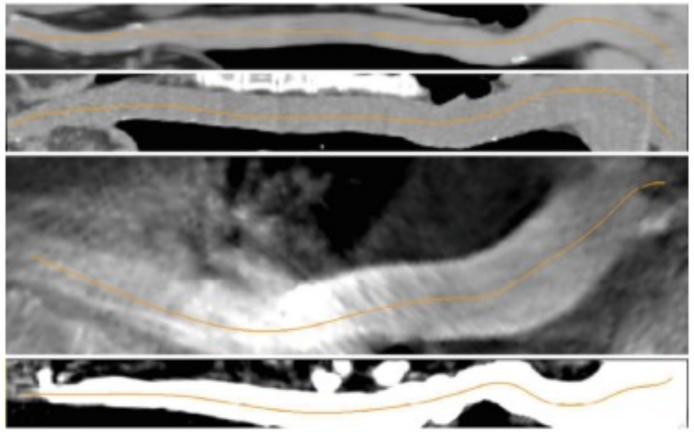

**Fig. 17. Example of traced aorta centerlines in the curved planar reformatting (CPR) view. Picture courtesy of Zhang et al. (2018).**

# DRL in Medical Imaging

- Solving Optimization using DRL

  - DRL is known to handle “non-differentiable metrics” therefore it is widely used for optimizations.

    1. Tuning Hyperparameter
    2. Image Augmentation
    3. Image Classification
    4. Searching Neural Network for Augmentation

# DRL in Medical Imaging

- Image Classification

  - CNN model integration with question and answering RL-Based method.
  - DNN agent asked the patient about the skin disease symptom using CNN visual information and answers.
  - DNN learned from this method based on question and answering the patient.
  - This approach improves the classification accuracy greater than 20% compared only to CNN using only visualization.
  - It shortened the diagnosis in average time compared to decision tree-based QA approach.

# DRL in Medical Imaging

- Image Segmentation
  - Assigning label to pixels helps in finding the perfect boundaries of anatomical structure of medical image.
  - This method realized the object image segmentation and mostly used for pre- processing. However, this method is intractable in size and not yet fully utilized to meet the clinical requirements for 3D image segmentation.
  - State of the art proposed NAS (Neural Architecture Search) improves the image search automation for special application however it is not used often in medical image segmentation.

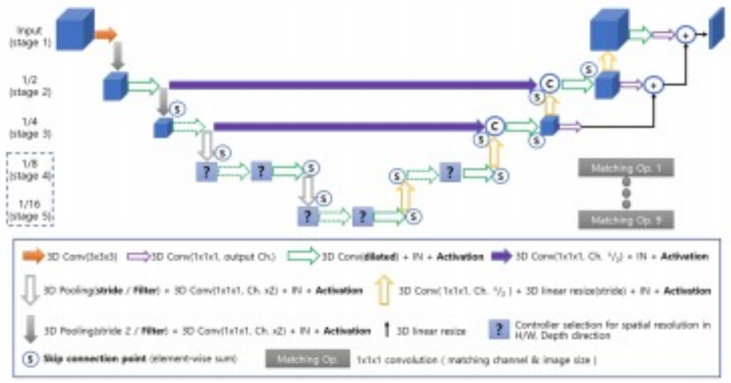

**Fig. 19. The proposed base architecture that is modified to best fit the application by using RL. Picture courtesy of Bae et al. (2019).**

# DRL in Medical Imaging

- Image Acquisition and Reconstruction

  - CT Metal artifacts affects the clinical decision making because of image inconsistency.

  - Using the iterative CT reconstruction helps to solve the optimization issue in medical image.

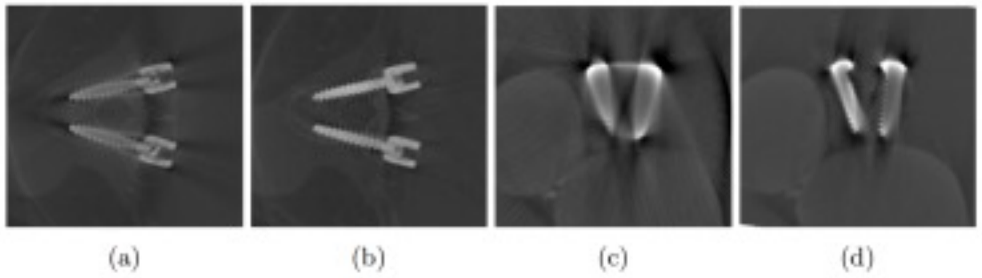


**Fig. 20. Two examples of axial slices from a volume reconstructed from (a,c) a straightforward short-scan and (b,d) a task-aware trajectory recommended by the agent. It is evident that the visual quality of the images reconstructed by using the agent is better. Picture courtesy of Zaech et al. (2019).**

# DRL in Medical Imaging

- Radiotheraphy Planning

  - DRL weight tuning policy was leverage in Radiotheraphy planning that takes the volume histogram as input and outputs the adjusted weight with reward function sparing the organs at risk.

  - This method proposed to improve the radiotherapy planning with yields of quality score 10.7% higher than human radiotherapy planner.

# DRL in Medical Imaging

- Miscellaneous Topics

  - These topics  are for exploration, not related to medical images analysis but mostly uses based Reinforcement Learning.

  1. Video Summarization
  2. Surgical gesture segmentation and classification
  3. Personalized mobile health intervention
  4. Computational model personalization

# Conclusion

- DRL is powerful model in medical imaging analysis.
- Successfully applied to various applications in landmark localization, object detection, image registration and image-based inferencing.
- Demonstrated effectiveness in tuning parameter optimization, image augmentation selection, and Neural Architecture Search (NAS).
- Although some methods are not fully utilized in clinical requirement the current DRL model surpassed my expectations and still needs further exploration in medical imaging classification.

Thank You!!!