Biomedical Physics & Engineering Express

OPEN ACCESS

RECEIVED
20 April 2023

REVISED
30 July 2023

ACCEPTED FOR PUBLICATION
18 August 2023

PUBLISHED
4 January 2024

PAPER

Learning to see via epiretinal implant stimulation *in silico* with model-based deep reinforcement learning

Jacob Lavoie , Marwan Besrour, William Lemaire , Jean Rouat , Réjean Fontaine and Eric Plourde

Department of Electrical Engineering and Computer Engineering, Université de Sherbrooke, Sherbrooke, Quebec, J1K 2R1, Canada

E-mail: jacob.lavoie@usherbrooke.ca

Keywords: deep reinforcement learning, model-based, retinal implant, stroke-based rendering

Supplementary material for this article is available online

Abstract

Objective: Diseases such as age-related macular degeneration and retinitis pigmentosa cause the degradation of the photoreceptor layer. One approach to restore vision is to electrically stimulate the surviving retinal ganglion cells with a microelectrode array such as epiretinal implants. Epiretinal implants are known to generate visible anisotropic shapes elongated along the axon fascicles of neighboring retinal ganglion cells. Recent work has demonstrated that to obtain isotropic pixel-like shapes, it is possible to map axon fascicles and avoid stimulating them by inactivating electrodes or lowering stimulation current levels. Avoiding axon fascicule stimulation aims to remove brushstroke-like shapes in favor of a more reduced set of pixel-like shapes. Approach: In this study, we propose the use of isotropic and anisotropic shapes to render intelligible images on the retina of a virtual patient in a reinforcement learning environment named rlretina. The environment formalizes the task as using brushstrokes in a stroke-based rendering task. Main Results: We train a deep reinforcement learning agent that learns to assemble isotropic and anisotropic shapes to form an image. We investigate which error-based or perception-based metrics are adequate to reward the agent. The agent is trained in a model-based data generation fashion using the psychophysically validated axon map model to render images as perceived by different virtual patients. We show that the agent can generate more intelligible images compared to the naive method in different virtual patients. Significance: This work shares a new way to address epiretinal stimulation that constitutes a first step towards improving visual acuity in artificially-restored vision using anisotropic phosphenes.

1. Introduction

Vision loss has a serious impact on quality of life. Epidemiological studies reveal that vision loss also has an important global burden of disease on society [1]. There are effective treatments for common eye diseases such as myopia, glaucoma, and cataracts. However, treatments aimed at prevalent diseases affecting the retina, such as age-related macular degeneration and retinitis pigmentosa, can slow the disease at best. Age-related macular degeneration accounts for 15.85% of incurable vision loss cases [1]. Therefore, it is the most prevalent untreatable disease that causes vision loss in developed countries [1]. Retinitis pigmentosa is a rare disease with a worldwide prevalence of 1/4000 that causes vision impairment to complete loss during adolescence and young adult life [2]. The early onset of retinitis pigmentosa increases the detrimental burden of the disease and remains one of the leading causes of blindness in the 20-year-old to 64-year-old age group [3, 4].

Age-related macular degeneration and retinitis pigmentosa cause degeneration of the retina's photoreceptor layer. Therefore, patients gradually lose their sensitivity to light, leaving subsequent layers of neurons with an aberrant signal. One treatment is to electrically stimulate the surviving neurons to artificially restore a certain visual acuity. This can be performed with the use of microelectrode arrays (MEA) that can be implanted to target different layers of the retina [5]. These devices were developed based on the observation that focal electrical stimulation of the retina generates a dot-shaped visual perception called phosphene [6]. Phosphenes are spatially preserved

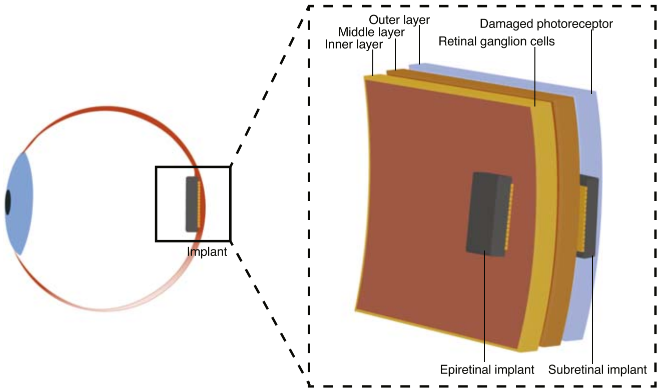


Figure 1. Epiretinal and subretinal implantation sites are illustrated in the anatomical context of the retina.

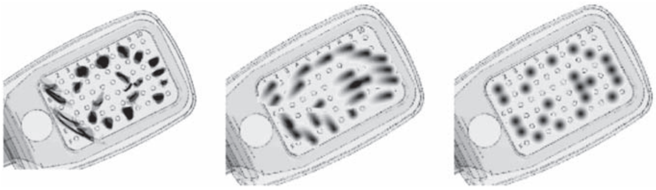


Figure 2. Comparison between anisotropic and isotropic phosphenes shapes. (Left) Subject 2 drawings from [12] show anisotropic phosphenes perceived during single-electrode stimulation with an Argus II implant. (Center) Axon map model predicts the anisotropic shaped caused by extracellular axon stimulation. (Right) Scoreboard model does not include extracellular axon stimulation resulting in ideal isotropic phosphenes. Phosphenes and images are generated using the *pulse2percept* Python module [14].

along the visual pathway as a result of the retinotopic organization of the visual system. An ensemble of phosphenes caused by retinal stimulation is referred to as a percept. The main focus of the work presented in this article is to train a reinforcement learning (RL) agent that selects phosphenes to generate a percept similar to a digital image in different virtual patients, thus restoring visual acuity. Before presenting the proposed approach, we go through the train of thought that led to our attempt to leverage anisotropic phosphenes instead of mitigating them.

1.1. Stimulation sites

There are two MEA implantation sites, shown in figure 1, which are often found in the literature [5, 7]: the subretinal and epiretinal implantation sites. In addition to the aforementioned sites, suprachoroidal, lateral geniculate nucleus, and visual cortex implants are also potential stimulation sites that target different parts of the visual pathway [8]. The epiretinal implantation site stimulates retinal ganglion cells (RGC), while the subretinal implantation site stimulates bipolar cells. Subretinal implants hold great promise in terms of visual acuity as a result of the lower-level signal encoded by bipolar neurons compared to subsequent layers [9]. However, the insertion point makes its deployment more complex. In fact, to be installed, the subretinal implant must be surgically implanted between the pigment epithelium and the outer retina. Therefore, its use is limited to patients with intact inner and middle layers of the retina [10]. In addition, data and power are generally transmitted through wires and an induction coil to the subretinal implant. Epiretinal implants can be placed on the retinal surface, allowing the use of optical power and data transmission [7, 11]. In this study, we focus mainly on epiretinal implant stimulation due to its potential to help more patients and be less invasive.

1.2. Psychophysical study of anisotropic phosphenes in epiretinal stimulation

Recent experiments with patients using an epiretinal implant revealed that phosphenes are not always dot-shaped [12]. In fact, although the epiretinal implant aims to stimulate the RGCs, it also stimulates the peripheral axons of the RGCs that are sufficiently close to the stimulating electrode. The stimulation of peripheral RGC axons causes the patient's visual system to render irregular shapes as shown in figure 2 [12, 13]. Axons are organized in fascicles, also known as axon bundles, which converge to the blind spot to form the optic nerve. Epiretinal electrical stimulation, therefore, generates an elongated shape parallel to the axon fascicles [12] as shown in figure 2. The phosphene is elongated along the axons, causing a perfect isotropic phosphene to become anisotropic.

1.3. Naive stimulation algorithm

The Naive Stimulation Algorithm (NSA) for the epiretinal implant described in [15] is tested in patients with the Argus II epiretinal implant. Many improvements to the original algorithms that use more advanced image processing methods such as Difference of Gaussian [16], constrained optimization [17], and deep learning [18, 19] obtained better results in simulation [20]. The NSA algorithm for epiretinal implant most tested in patients is transforming a digital image into electrical stimulation using a downscaling operation of a camera image that matches the intensity of a pixel to the amplitude or frequency. This results in a pixelated image having the dimension of the microelectrode array. Electrical stimulations are proportional to the intensity of the pixels. These electrical stimulations are delivered in a temporal sequence of single-electrode stimulations in an experimental setup preventing eye movement. Simultaneous stimulation with multiple electrodes causes more irregular phosphene shapes [21]. Therefore, the NSA stimulation algorithm does not consider the fact that RGC axons are stimulated, leading to anisotropic phosphene [15].

1.4. Mitigating axon bundle stimulation

As indicated above, axon bundle stimulation produces anisotropic phosphenes. Previous work attempted to minimize the impact of anisotropic shapes on the percept quality. One such approach consists of modifying the electrode configuration to attenuate this impact [22, 23]. Other approaches adopt different stimulation techniques, such as RGC mapping and current steering, creating virtual electrodes between electrodes, resulting in more consistent isotropic phosphenes with a healthy retina *in situ* [24–31]. These approaches offer better control over RGC spiking and collateral axon stimulation. In addition, both tend to limit the number of usable electrodes and the range of possible stimulation intensity, thus reducing the diversity of shapes that can be generated by epiretinal implants [23, 25].

1.5. Leveraging axons bundle stimulation with stroke-based rendering

Data acquired from patients with epiretinal implants and anatomical studies allow the development of a psychophysically validated model of end-to-end visual processing in the degenerated retina [12]. These models help to visualize the perceived anisotropic shapes created by stimulation of axon bundles in an epiretinal stimulation setting, as shown in figure 2.

Instead of designing a retinal stimulation algorithm that mitigates anisotropic shapes, it is possible to use the available shapes produced by all possible stimulations to form the desired image to be perceived. Choosing anisotropic shapes to generate an image is referred to in the computer vision community as stroke-based rendering (SBR). More precisely, SBR is a non-photorealistic method to create imagery from discrete elements called strokes, such as paint strokes or ripples [32]. An analogy that illustrates the problem is that of the painter reproducing a photograph on a canvas. In this paper, the phosphene is considered equivalent to one brushstroke, and the retina is the canvas.

1.6. Reinforcement learning in stroke-based rendering

Recent successful SBR algorithms such as SPIRAL [33], StrokeNet [34], Doodle-SDQ [35], Sketch-RNN [36] and other model-based approaches [37] showed high-quality picture reproduction. All successful attempts mentioned the use of deep reinforcement learning (DRL) paradigm for paint stroke decomposition and/or Generative Adversarial Networks (GAN) for quality assessment. Thinking of epiretinal stimulation as an SBR problem allows us to train a deep reinforcement learning agent to use the full diversity of shapes produced by all electrodes. This new perspective offers the use of anisotropic shapes rather than mitigating them as in the aforementioned approaches in section 1.4.

1.7. Reinforcement learning and retinal stimulation

Previous work using reinforcement learning (RL) to adjust epiretinal stimulation parameters used the center-surround RGC receptive field model as a premise to simulate the retina [38, 39]. The center-surround RGC receptive field is the accepted model for RGC neural coding. It consists of a circular area of the retina called the center and the surrounding region that respond oppositely to light exposure. However, it does not consider the stimulation of axon bundles in the context of epiretinal stimulation. Some more recent work [40] using DRL is promising but does not include anisotropic phosphene in the percept generation. Nonetheless, as the author notes in [38], RL is a very appealing framework to use patient's feedback as a learning signal to automate adaptation to different patients *in vivo* [41].

For each patient, the best ensemble of phosphenes that render a particular target image on the retina is not known. The current virtual patient models mentioned in figure 2 can only generate a percept from electrode stimulation. One way to find the best electrode stimulation combination from a target image is to evaluate the perception of the predicted percept compared to the target image. A brute-force approach could be to try every combination of electrodes, generate the percept, and then calculate the similarity between the percept and the target image. This approach becomes more tedious as the state-space, or in this case, the number of electrodes increases in complexity [42]. Therefore, it is difficult to generate a complete dataset to leverage supervised learning methods. RL allows for searching the state-space of electrode combinations more efficiently [42]. A DRL agent that interacts with an environment receives

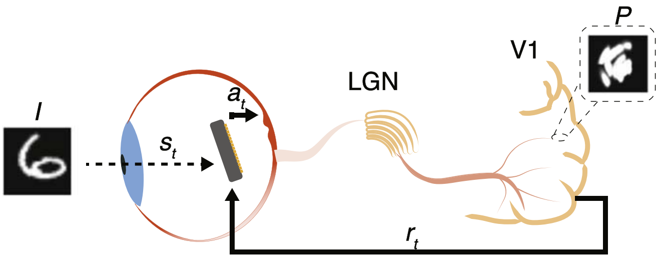


Figure 3. A model-free environment simulates a virtual patient's retina and its visual system. The agent is defined as the implant controller and is in charge of transforming a camera image into a single-electrode stimulation pattern. The single-electrode stimulation excites the retina. The signal propagates to the lateral geniculate nucleus (LGN) and through the visual cortex (V1) and subsequent visual areas (Not illustrated). The environment returns a state s_t. The agent must interact with the environment through an action a_t corresponding here to the electrode's stimulation. The environment then gives a reward r_t according to the similarity between the target image I and the visual perception P resulting from the electrical stimulation.

direct feedback through a reward as it approaches a solution. The agent learns to associate a particular state of the environment with an action it can take. RL is therefore a suitable paradigm for epiretinal stimulation because there exists a model of epiretinal stimulation with which an agent can interact, but there is no optimal solution to transform a target image into an electrode combination.

1.8. Proposed approach

This work bridges the gap between the previous attempt [38] to find optimal stimulation parameters with an epiretinal implant using RL and the latest anatomical knowledge and lessons learned from trials of human epiretinal implants [12]. It is the only attempt to improve visual acuity for the implanted patient by leveraging anisotropic phosphene in a SBR problem.

More specifically, we want to investigate whether a DRL agent can learn to generate a sequence of single-electrode stimulation from a target image, thus increasing visual acuity for different virtual patients with an epiretinal implant. This paper proposes the following contributions:

- We formalize the transformation of the original image into a stimulation pattern as an SBR problem implemented in a new reinforcement learning environment named *rlretina* available at https://github.com/NECOTIS/rlretina.git.
- We compare a pixel-based distance and distance between probability distributions as a proxy of a perceptual metric in the reward design of the new environment.
- We build a model-based DRL agent that can explore the available anisotropic phosphene space to increase visual acuity in virtual patients with different epiretinal implant settings.
- Finally, we investigate perceptual metrics to circumvent pixel-based metric limitations to better compare the percepts resulting from different algorithms.

In addition, we demonstrate pixel-based error limitations as a reward in this perceptual task. We compare the proposed DRL agent with the NSA stimulation algorithm using the mean structural similarity index measure (MSSIM) as an evaluation metric presented in section 2.2.2. We show that the proposed agent better reproduces the original images in the virtual patient's percepts than the NSA stimulation methods. It also performs better with different virtual patient implant settings than the NSA stimulation algorithm.

1.8.1. *Using a model-based approach*

A model-free RL approach to epiretinal stimulation in an *in vivo* system is presented in figure 3. The state s_t contains only the digital image I. It is the reward r_t that indirectly provides information about the similarity between I and the percept P. The virtual patient's visual system illustrated in figure 3 is conceptualized as part of the environment (see section 2.1.1). This approach was used in previous work presented in section 1.7.

In the present work, we adopt a model-based approach to data generation explained in section 2.1.2 by including P in s_t as shown later in figure 4. Using an external world model to generate the data for the agent, in this case, the percept P, is derived from the recurrent World Models proposed in [43]. It alleviates the agent's burden of modeling the environment that can be composed of any real-world image. Instead, it uses the axon map model to generate the image.

1.8.2. *Epiretinal stimulation as Markov Decision Process*

A Markov Decision Process (MDP) allows one to formalize the interaction between the agent and the environment in a RL task. The proposed approach is inspired by the way the SBR task is formalized as a MDP where the goal is to maximize the similarity between a target image and a stroke-based rendered image equivalent respectively to I and P in figure 3.

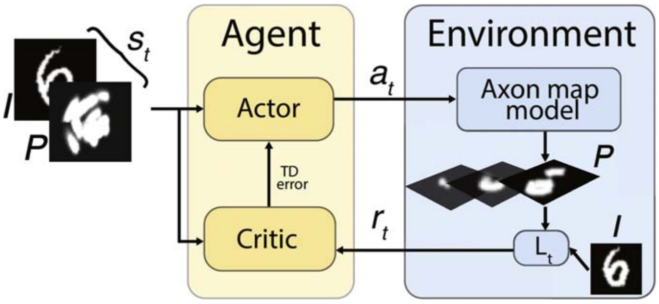


Figure 4. In contrast to figure 3, the SAC agent receives the percept P as well as the original image I. This model-based data generation scheme makes use of the predictive model of the retina to give access to the internal state L_t of the environment model. Electrical stimulations are transformed into a visual percept using the axon map model from *pulse2percept* library. The SAC agent is composed of an actor and a critic that takes I and P as input s_t. The actor is in charge of estimating the action a_t equivalent to a single-electrode stimulation. The critic gives evaluative feedback to the actor taking action a_t at state s_t.

Selecting single-electrode stimulation in a sequence is a decision-making task that can be modeled as a MDP. A finite MDP is a decision-making process that satisfies the Markov property stipulating that the action influences not only the immediate reward r_t, but also the probability that the process moves into its new state s' at $t+1$ [44]. A finite MDP is defined by sets with a finite number of elements for states $\mathcal{S}$, actions $\mathcal{A}$, and rewards $\mathcal{R}$. Given s_t the state at time t, and a_t the action at time t, the dynamic or state transition function of the environment p is defined in (1) [44].

$$p(s', r|s, a) \doteq Pr\{s_{t+1} = s', r_{t+1} = r|s_t, a_t\} \quad (1)$$

[for all s' where $s \in \mathcal{S}$, $r \in \mathcal{R}$ and $a \in \mathcal{A}(t)$. In other words, the next state s_{t+1} and the associated reward r_{t+1} are functions of the probability Pr of taking action a_t in the previous state s_t. The state s_t given to the agent is defined as the target image I, and the percept P that exposes the internal dynamics of the retina model. The agent must then decide on an action a_t, namely, a single-electrode stimulation. As mentioned previously, single-electrode stimulation generates anisotropic shapes perceived here by the virtual patient. The percepts of the virtual patient are simulated with a psychophysically validated model developed from a human patient implanted with an Argus epiretinal implant [12]. This model of human epiretinal vision is used by the environment that simulates a virtual patient.

2. Materials and methods

The following sections present the implant stimulation algorithm and the experiences that result in the generation of intelligible percepts. The algorithm takes the form of a DRL agent detailed in section 2.1.2 estimating the best sequence of action a_t to maximize the similarity between I and P to increase the visual acuity of the patient.

2.1. Materials

2.1.1. Environment

We developed a new environment to train reinforcement learning agents. The environment follows the OpenAI gym specifications [45]. Section 2.1.1 presents the underlying assumptions in the environment.

From single-electrode stimulation to percept The environment uses the axon map model that accurately reproduces the drawings of the phosphene perceived by real patients during an epiretinal single-electrode stimulation [12]. It is important to note that the axon map model was developed in a control clinical set-up that prevented eye movements and used only single-electrode stimulation without superimposition of multiple stimulations. Therefore, it serves as an anisotropic phosphene rendering tool to simulate the virtual patient. The axon map model, detailed in the section below, renders anisotropic phosphenes as observed in implanted patients rather than ideal isotropic phosphenes. Rendered phosphenes are assembled to form the percept P. In the case at hand, an episode of agent-environment interaction is defined as a sequence of single-electrode stimulations. At the beginning of the episode, a new target image I is selected from the image dataset. At each step or single-electrode stimulation of an episode, the agent produces a vector with probabilities of selecting each electrode. The number of activated electrodes and their normalized stimulation values are set in the environment configuration. The environment uses single-electrode stimulation equivalent to the action a of the agent to update the virtual patient's percept P.

Axon map model An electrical stimulation is assumed to generate focal dots of light that decay exponentially with the distance between the location of the stimulating electrode and the location of the stimulated retina (x_{stim}, y_{stim}) and the spatial decay constant ρ. These assumptions are included in the calculation of the scoreboard model to estimate the intensity profile $I_{score}(x, y; \rho)$ [12]. Equation (2) corresponds to the scoreboard model presented in figure 2 [12].

$$I_{score}(x,\, y;\, \rho) = \exp\left(-\frac{(x - x_{stim})^2 + (y - y_{stim})^2}{2\rho^2}\right) \quad (2)$$

The axon map model estimates the contribution of axon stimulation to the virtual patient's perception based on anatomic observation of axonal growth [46]. The axonal trajectories are represented with a modified polar coordinate system with its center at the optical disc. The contribution of axon stimulation to a phosphene decays exponentially along the axon bundles with the distance between the location of the stimulating electrode and the soma (x_{soma}, y_{soma}). The impact of axon stimulation on the intensity profile $I_{axon}(x, y;\, \lambda)$ is estimated using (3).

$$I_{axon}(x,\, y;\, \lambda) = \exp\left(-\frac{(x - x_{soma})^2 + (y - y_{soma})^2}{2\lambda^2}\right) \quad (3)$$

where λ is a constant that modulates spatial decay along the axon. Therefore, we can combine (2) and (3) to predict the intensity profile $I_{map}(x,\, y;\, \rho,\, \lambda)$ of anisotropic phosphenes perceived by virtual patients implanted with an epiretinal implant:

$$I_{map}(x,\, y;\, \rho,\, \lambda) = I_{score}(x,\, y;\, \rho) I_{axon}(x,\, y;\, \lambda) \quad (4)$$

The values of parameters ρ and λ can be set to simulate implant placement relative to the retina of different virtual patients, as demonstrated with real patients in [12].

Implant simulation A model of the commercialized ArgusII implant [15] is used in the simulations presented in the current work to facilitate the reproducibility of the experiments and further comparisons with other stimulation algorithms. The implant is placed on the surface of the retina according to the coordinates centered on the fovea. The angle of insertion is set through the environment configuration. The implant placement parameters of subjects in [12] are also available in the environment. All single-electrode pulse waveforms consist of a biphasic, cathodic-first, charge-balanced, square-wave pulse.

Reward definition The reward is defined as follows:

$$r(s_t,\, a_t) = \frac{L_{t+1} - L_t}{L_{t_0}} \quad (5)$$

where L_t is a given distance between the target image I and the percept P at time t and t_0 indicates the beginning of the experiment. The difference in two subsequent time steps, $L_{t+1} - L_t$, is used to signal the agent if it is getting closer or farther from the target.

Dataset The MNIST [47] dataset used in the experiments serves as a proxy for visual acuity tasks often used in optometry, such as the Snellen chart [48]. It contains 70 000 28 × 28 handwritten number images split into a 60 000 image training set and a 10 000 image test set.

Table 1. SAC hyperparameters.

Parameters	Values
Learning rate	$3 \cdot 10^{-4}$
Discount factor (γ)	0.99
Replay buffer	10^6
Number of hidden layers	3
Number of hidden units by layers	512
Number of samples by minibatch	32
Nonlinearity	ReLU
Reward scale	200
Target smooth coefficient (τ)	0.0005
Target update interval	1

2.1.2. Agent

Building of a model-based agent The agent architecture is the Soft Actor-Critic (SAC) algorithm [49]. An actor-critic agent is made up of two parts; (1) the actor who learns a policy $\pi(a_t|s_t)$ that maps a state $s_t \in \mathcal{S}$ to actions a and (2) the critic who approximates a value function that gives an evaluative feedback based on the agent's action a_t in state s_t. The SAC algorithm uses a Q-function in the critic in a similar way to recent agent algorithms such as the Deep Deterministic Policy Gradient (DDPG)[50, 51]. This allows the agent to be trained in an off-policy manner and, therefore, reuse data efficiently compared to the standard policy iteration used in the classic actor-critic formulation [44]. The SAC algorithm uses a stochastic actor and maximizes the entropy of the actor with an entropy maximization objective. This results in a more stable and scalable algorithm that exceeds the efficiency and final performance of DDPG [49]. The associated SAC cost function is as follows :

$$J(\pi) = \sum_{t=0}^{T} \mathbb{E}(s_t,\, a_t)_{\sim\rho_\pi}[r(s_t,\, a_t) + \alpha H(\pi(a_t|s_t))] \quad (6)$$

where the parameter α controls the stochasticity of the optimal policy during training and $\mathbb{E}$ denotes the mathematical expectation [49]. Like in standard RL, the SAC algorithm maximizes the expected sum of rewards $\sum_{t=0}^{T}\mathbb{E}(s_t,\, a_t)_{\sim\rho_\pi}[r(s_t,\, a_t)]$. As mentioned above, the SAC algorithm also includes the expected entropy $\sum_{t=0}^{T}\mathbb{E}(s_t,\, a_t)_{\sim\rho_\pi}[H(\pi(|s_t))]$ of the policy over $\rho_\pi(s_t)$ in the loss function $J(\pi)$ as shown in (6). In the proposed approach, both the actor and the critic are approximated using a convolutional neural network (CNN) with a CoordConv [52] layer (See table 1).

As illustrated in figure 4, the actor and the critic take the state s_t composed of the target image I and the percept P at time t. The agent does not need to model the environment implicitly, as opposed to a model-free approach. The transition dynamic of the environment corresponds to the axon map model that generates the percept P. The percept is then given directly to the agent in s_t. Therefore, the agent uses model-based data generation.

Training of the agent This section presents the details of the agent's training implemented with the *Ray RLlib* deep reinforcement learning library [53]. All hyperparameters to replicate the agent are collected in table 1. The agent interacts with the environment until it reaches the N steps corresponding to N single-electrode stimulation. N steps form an episode. The agent is trained *tabula rasa* with batches of 32 episodes over 1000 iterations. The agent's replay memory buffer that allows for more stable learning and off-policy training is set to hold the latest 10^6 steps [54]. The Adam [55] optimizer is used to train the neural networks of the actor and the critic. The agent is trained on two AMD Milan 7413 with 24 cores running at 2.65 GHz and one NVidia A100 GPU.

2.2. Methods

To validate the reward design of the environment and compare the agent with the NSA approach, two experiments were carried out (1) a pivotal experiment using pixel-based and perception-based metrics as reward to obtain a readable percept produced by the agent and (2) an experiment that demonstrates the agent's ability to adapt to different patients.

2.2.1. Effect of reward shaping on agent's learning

Having a suitable metric to measure the pixel and perceptual similarities between the percept and the target image is critical to the agent's training. The reward defined in (5) gives the agent a clear signal of whether it is getting closer to the target image faster or slower. We compared both l_2 and Wasserstein distances to estimate L_t in (5) in the hope of accelerating learning with better reward shaping [56]. l_2 distance is used as a reference metric similar to the pixel-based metrics commonly used in computer vision. Wasserstein distance is a probability distribution-based metric. It is an estimate of the distance between two probability distributions. The Wasserstein distance is estimated using the Sinkhorn iteration algorithm from a maximum entropy perspective between the two probability distributions [57].

Optimizing for pixel-error such as l_2 distance encourages finding pixel-wise averages for a plausible solution. It typically results in the loss of high-frequency details, giving overly smooth images with poor perceptual quality [58, 59]. Wasserstein distance allows one to better preserve the probability distribution of the light in the image [56]. Performances of the two metrics in training a DRL agent in the environment are presented in section 3.1.

2.2.2. Percept quality in different virtual patients

Pixel-based versus perceptual-based comparison Agent-generated percepts are compared to the NSA algorithm described in section 1.3. The images are resized according to the Argus II layout of 10 by 6 electrodes. The intensity of the pixels is uniformly assigned to the amplitude of the electrical pulse described in section 2.1.1.

We use the l_2 norm and the mean squared error (MSE), which are two *de facto* standard in image restoration [60] to compare percepts generated by the agent and the NSA algorithm. We define the metrics for M by N images as follows:

$$l_2 = \sum_{i=0,j=0}^{M,N} \sqrt{(I_{i,j} - P_{i,j})^2} \tag{7}$$

$$MSE = \sum_{i=0,j=0}^{M,N} \frac{(I_{i,j} - P_{i,j})^2}{MN} \tag{8}$$

However, they do not correlate well with image quality as perceived by the human visual system [60]. Therefore, we use the mean structural similarity index measure (MSSIM), which is a perceptually motivated metric [61]. We calculate the MSSIM with non-negative image patches $\mathbf{x}$ and $\mathbf{y}$ of size 7 by 7 as follows:

$$MSSIM(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_{\mathbf{x}}\mu_{\mathbf{y}} + c_1)(2\sigma_{xy} + c_2)}{(\mu_{\mathbf{x}}^2 + \mu_{\mathbf{y}}^2 + c_1)(\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{y}}^2 + c_2)} \tag{9}$$

$c_1 = 0.01$ and $c_2 = 0.03$ are small constants that add numerical stability when the means μ or the standard deviation σ are close to zero [61]. MSSIM takes into account the local characteristics of the image in a way similar to that of the human visual system. On the contrary, pixel-based metrics, such as the l_2 norm and MSE, evaluate the difference between the corresponding pixels of two images independently of the nearby pixels.

Adaptation to different virtual patients Two SAC agents are trained on two virtual patients with different ρ constant and N single-electrode stimulation. The spatial decay constant ρ is varied to simulate two realistic virtual patients with different distances between the electrodes and the retina. To ensure that a phosphene is anisotropic, its shape must be dominated by axon fascicle stimulation ($\lambda > \rho$). The number of steps in an episode equivalent to the number of single-electrode stimulation N also increases. The high ρ and high N make the task more difficult for both approaches because the algorithms must deal with many large phosphenes.

3. Results

3.1. Effect of reward shaping on agent's learning

Figure figure 5 shows that the reward estimated by the l_2 distance quickly saturates to a reward value after only 1000 episodes, while the reward estimated by the Wasserstein distance slowly progresses. The results with a conventional l_2 distance as an estimator of L_t are very limited. Looking at the samples as shown in figure 5 reveals that the l_2 distance only taught the agent to represent low spatial frequencies of the dataset, while the Wasserstein distance preserves the particularity of each character. The agent's reward then saturates as shown in figure 5 since it fails to learn to use the finer structures of the image to increase its reward. The behavior persists in experiments (not presented here) that include data augmentation in an

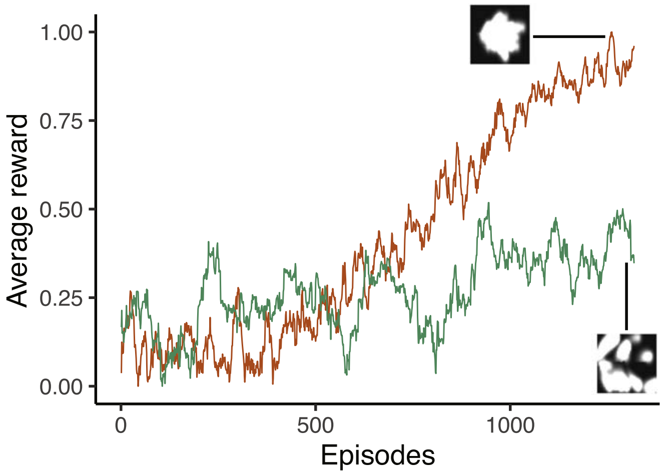


Figure 5. Comparison of the episode average reward of two SAC agents during training using respectively l_2 (red) and Wasserstein (green) based reward obtained by the SAC agent. Representative samples of percept produced by the agents early during training are shown. The reward average are normalized over a training iteration for the comparison.

Table 2. Metric evaluating the percept and target image. Only the experiment in which NSA obtains readable percept ($\rho = 200$ and $\lambda = 500$) is shown. Lower l_2 and MSE is better. A higher MSSIM is better. The mean and standard deviation are calculated on 1000 MNIST images.

Metric	Random	NSA [15]	SAC Agent
l_2 norm	13.54(1.67)	**11.69(1.63)**	12.58(2.07)
MSE	0.11(0.02)	**0.08(0.02)**	0.09(0.03)
MSSIM	0.07(0.05)	0.28(0.08)	**0.35(0.11)**

effort to remove bias from the MNIST dataset toward high-luminance in the center.

A Wasserstein-based reward stabilizes learning and allows more exploration without catastrophic failures as opposed to the l_2-based reward. Wasserstein-based reward gives useful positional information regarding the distance of the phosphene from the high-luminance region of the image. l_2 distance fails to give this information through the reward since it is calculated pixel-wise. Therefore, the use of a Wasserstein-based reward results in more persistent electrode stimulation outside the high-luminance region. Random single-electrode stimulation is added to table 2 to better understand the results, as no other reference can be used as a benchmark. Random stimulation helps to better grasp the range of values specific to each metric in the environment. It also helps to contextualize pixel-based metrics with a tighter range, such as the l_2 norm and MSE, with perceptual metrics such as MSSIM.

3.2. Percept quality in different virtual patients

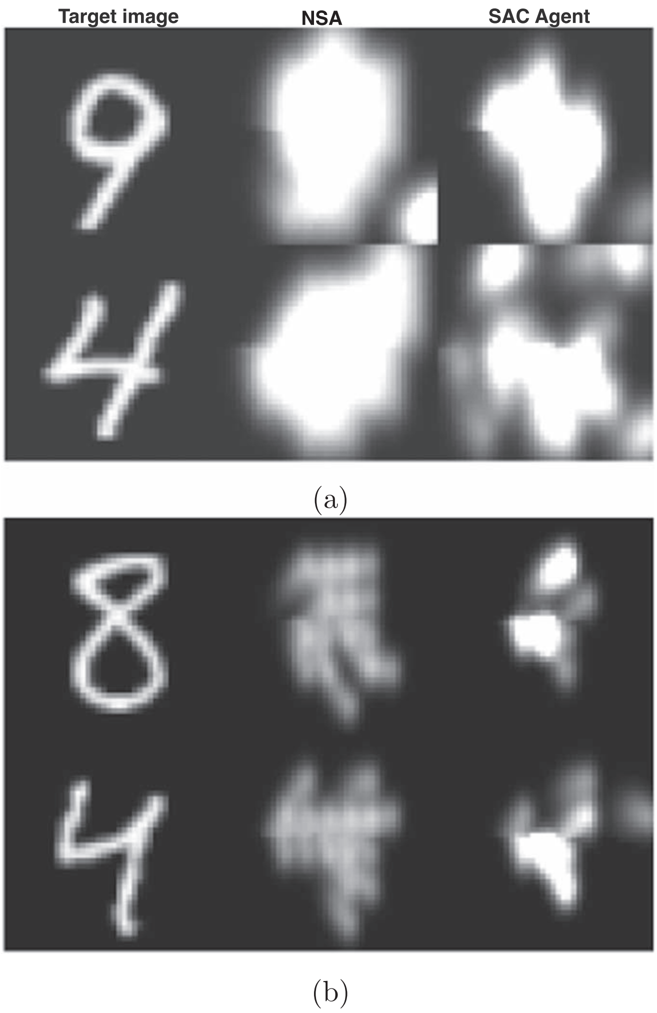
The experiment presented in the previous section establishes that the best metric to evaluate the similarity between I and P is the Wasserstein distance as an estimator L_t. Therefore, the percepts generated by the SAC agent in figure 6 are obtained by training the agent with the Wasserstein distance. In section 3.2.2, the agent is trained in different virtual patients to demonstrate the flexibility of the proposed approach over the NSA stimulation algorithm.

3.2.1. Pixel-based versus perceptual-based comparison

Table 2 shows the image reconstruction metrics for the $\rho = 200$ and $\lambda = 500$ experiment in which the implant is close enough to the retina to obtain intelligible percepts with the NSA algorithm. Samples of the percept generated for this virtual patient are shown in figure 6(a). NSA single-electrode stimulations are, by default, limited to the high-luminance region of the image. However, the SAC agent has the freedom to choose any single-electrode stimulation. This is particularly visible when comparing the NSA and SAC agent samples in the second rows of figures 6(a) and (b). This results in single-electrode stimulation in the out-of-high luminance region, thus increasing the l_2 distance and MSE for the SAC agent (See table 2). However, the SAC agent significantly outperforms the NSA algorithm in preserving the structural integrity of the image measured with MSSIM ($t = 15.71$, $p < 0.001$) as observed in figure 6. NSA has significantly lower error-based metrics than the SAC agent (l_2; $t = 10.66$, $p < 0.001$ and MSE: $t = 10.77$, $p < 0.001$). The t-test are computed with a sample of 1000 MNIST images. It shows that an SAC agent, despite higher values in error-based metrics, increases the readability of the digits, in contrast to the NSA algorithm.

3.2.2. Adaptation to different virtual patients

These experiments were carried out to demonstrate the limits of NSA algorithms in the condition often observed in implanted patients where the implant is relatively far from the retina [12]. A high distance between the stimulating electrode and the retina (or a high ρ) produces larger phosphenes that result in low



(a)

(b)

Figure 6. Two virtual patient's percepts with different axon map model and single-electrode stimulation parameters. Columns from left to right are respectively the initial MNIST target image, the NSA generated percept and the SAC agent generated percept. (a) Samples generated with $\rho = 315$, $\lambda = 500$ and $N = 32$. (b) Samples generated with $\rho = 200$, $\lambda = 500$ and $N = 16$. Axon map model parameters are influenced by experiments realised with implanted subjects. Single-electrode stimulation is a biphasic pulse with a fixed amplitude of 10 μA.

resolution [62], as shown in figure 6(b). The NSA generated percepts are only readable with the implant close to the retina (low ρ) and with fewer single-electrode stimulations (low N), as observed by comparing figure 6(a) and figure 6(b). SAC agent better preserves the readability of characters in figure 6(b) when the implant is further from the retina (high ρ) and a large number of single-electrode stimulations (high N). Therefore, the SAC agent can better adapt to more restrictive implant placements in patients in terms of resolution.

4. Discussion

In this paper, we propose to address the anisotropic phosphene problem in epiretinal stimulation as an SBR task. We present a DRL agent that can improve visual perception of numbers in the context of epiretinal stimulation.

We emphasize the fact that the reward design of the new environment and the evaluation of the agent's performances must take into account the perceptual nature of the task. We address this limitation found in the reward design by comparing the l_2 and Wasserstein distances to assess the differences between the percept and the target image. The agent rewarded with the l_2 distance as an estimator uses only the electrodes near the center of the image, resulting in indistinguishable characters. A Wasserstein-based reward thus outperforms the l_2-based reward, as it conveys information about the distance between the distribution of light of I and P rather than the pixel-sharp l_2-based reward. This phenomenon is similar to the mode collapse phenomenon observed in GAN [63]. Using a perceptual-based metric as a reward, such as Wasserstein, allows learning to stabilize and avoid mode collapse observed with a pixel-based metric. More research is needed to assess whether the Wasserstein-based distance eliminates this phenomenon, as does the Wasserstein GAN algorithm [64].

The limitation of pixel-based metrics to evaluate the agent's performance is solved by proposing the more perceptually relevant MSSIM index. These results align with other recent SBR approaches that account for the human perception instead of pixel-wise error metrics in the loss function [56, 58, 59]. This is a first attempt to introduce human perception metrics in the design of a DRL stimulation algorithm for epiretinal implants. Future work aimed at improving and restoring vision in patients should use metrics that account for human perception instead of methods that use pixel-based accuracy.

The SAC agent outperforms the NSA algorithm in preserving the similarity between target images and the generated percepts in different virtual patients. It shows that the SAC agent adapts to different difficulty levels of the task as opposed to the NSA algorithm. This is important in the context of the real patient. It has been shown to be difficult to simultaneously modulate the size and brightness of phosphene [65]. The SAC agent allows one to circumvent certain limitations of the NSA algorithm, such as decreasing the intensity or inactivating electrodes in the case of anisotropic phosphenes or too large phosphenes. Moreover, it automatically learns an optimal single-electrode stimulation sequence without human percept inspection and tuning. It only requires standard hyperparameter tuning of the DRL agent. However, previous conclusions are limited by the fact that single-electrode stimulations are assumed to be independent of each other based on the protocol used to develop the axon map model [12].

Limitations regarding the speed of the implementation of the axon map model during these experiments considerably slowed data generation. Recent work [66] by the authors of the axon map model offers hope to replace the current implementation with a neural network. This approach uses an approximation of the axon map model with a neural network which allows experiments to use only a GPU, eliminating the need for data transfer between the CPU and GPU memory. This could significantly increase the percept generation process and the calculation of the reward in the proposed environment. Therefore, accelerating the agent training process.

The training procedure could be further improved with imitation learning. The agent currently learns from its interaction with the environment. Using a dataset generated with the NSA algorithm could serve as a baseline training before training directly in the environment. This could potentially improve the out-of-high luminance stimulation observed in the samples presented in figure 6.

5. Conclusion

In summary, this paper demonstrates that the formalization of epiretinal stimulation as an SBR problem allows for the full diversity of anisotropic phosphenes to be exploited, as opposed to the current NSA approach. Previously unwanted phosphene shapes can now expand the complexity of possible percepts for patients with an epiretinal implant. Further studies introducing metrics based on human perception into algorithm design could enhance the quality of recovered vision. This allows us to better personalize the algorithm for each patient and create a better user experience. This opens new ways to significantly improve the visual acuity of patients implanted with an epiretinal implant.

Acknowledgments

The support of the Digital Research Alliance of Canada and NSERC are gratefully acknowledged.

Data availability statement

The data that support the findings of this study will be openly available following an embargo at the following URL/DOI: https://github.com/NECOTIS/rlretina.

Declaration of competing interest

The authors declare that they have no competing interests.

ORCID iDs

Jacob Lavoie https://orcid.org/0000-0003-3741-4701
William Lemaire https://orcid.org/0000-0002-1780-1395
Jean Rouat https://orcid.org/0000-0002-9306-426X
Réjean Fontaine https://orcid.org/0000-0002-5453-0168

References

[1] Olusanya B O *et al* 2018 Developmental disabilities among children younger than 5 years in 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016 *The Lancet Global Health* **6** e1100–21

[2] Hartong D T, Berson E L and Dryja T P 2006 Retinitis pigmentosa *The Lancet* **368** 1795–809

[3] Buch H, Vinding T, La Cour M, Appleyard M, Jensen G B and Nielsen N V 2004 Prevalence and causes of visual impairment and blindness among 9980 Scandinavian adults: the Copenhagen City Eye Study *Ophthalmology* **111** 53–61

[4] Klaver C C W, Wolfs R C W, Vingerling J R, Hofman A and de Jong P T V M 1998 Age-Specific Prevalence and Causes of Blindness and Visual Impairment in an Older Population: The Rotterdam Study *Archives of Ophthalmology* **116** 653–8

[5] Chuang A T, Margo C E and Greenberg P B 2014 Retinal implants: a systematic review *British Journal of Ophthalmology* **98** 852–6

[6] Humayun M S, De Juan E, Dagnelie G, Greenberg R J, Propst R H and Phillips D H 1996 Visual perception elicited by electrical stimulation of retina in blind humans *Archives of ophthalmology* **114** 40–6

[7] Zrenner E 2002 Will retinal implants restore vision? *Science* **295** 1022–5

[8] Kleinlogel S, Vogl C, Jeschke M, Neef J and Moser T 2020 Emerging approaches for restoration of hearing and vision *Physiol. Rev.* **100** 1467–525

[9] Palanker D, Le Mer Y, Mohand-Said S and Sahel J A 2022 Simultaneous perception of prosthetic and natural vision in AMD patients *Nat. Commun.* **13** 1–6

[10] Pavlova P, Boneva J and Shandurkov I 2019 Epiretinal vs. subretinal implant in surgical treatment of retinitis pigmentosa-a review *Bulgarian Review of Ophthalmology* **63** 13–8

[11] Lemaire W *et al* 2021 Retinal Stimulator ASIC Architecture Based on a Joint Power and Data Optical Link *IEEE J. Solid-State Circuits* **56** 2158–70

[12] Beyeler M, Nanduri D, Weiland J D, Rokem A, Boynton G M and Fine I 2019 A model of ganglion axon pathways accounts for percepts elicited by retinal implants *Sci. Rep* 1–16

[13] Tsai D, Chen S, Protti D A, Morley J W, Suaning G J and Lovell N H 2012 Responses of retinal ganglion cells to extracellular electrical stimulation, from single cell to population: model-based analysis *PLoS One* **7** e53357

[14] Beyeler M, Boynton G M, Fine I and Rokem A 2017 *pulse2percept: A Python-based simulation framework for bionic vision Proc. of the 16th Python in Science Conf.* 81–8

[15] Luo Y H L and Da Cruz L 2016 The Argus® II retinal prosthesis system *Progress in Retinal and Eye Research* **50** 89–107

[16] Guo F, Duan J, Huang S, Xiao Y and Chu R 2023 Edge detection algorithm based on difference of gaussian for visual prosthesis *2023 IEEE 6th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), IEEE* vol 6 1331–4

[17] Spencer M J, Kameneva T, Grayden D B, Meffin H and Burkitt A N 2019 Global activity shaping strategies for a retinal implant *J. Neural Eng.* **16** 026008

[18] Wu Y, Karetic I, Stegmaier J, Walter P and Merhof D 2023 A deep learning-based in silico framework for optimization on retinal prosthetic stimulation arXiv:2302.03570

[19] Granley J, Relic L and Beyeler M 2022 Hybrid neural autoencoders for stimulus encoding in visual and other sensory neuroprostheses *Advances in Neural Information Processing Systems* 35

[20] Borda E and Ghezzi D 2022 Advances in visual prostheses: engineering and biological challenges *Progress in Biomedical Engineering* **4** 032003

[21] Rizzo J F, Wyatt J, Loewenstein J, Kelly S and Shire D 2003 Perceptual efficacy of electrical stimulation of human retina with a microelectrode array during short-term surgical trials *Investigative Ophthalmol. Vis. Sci* **44** 5362–9

[22] Esler T B, Kerr R R, Tahayori B, Grayden D B, Meffin H and Burkitt A N 2018 Minimizing activation of overlying axons with epiretinal stimulation: the role of fiber orientation and electrode configuration *PLoS One* **13** e0193598

[23] Bruce A and Beyeler M 2022 Greedy optimization of electrode arrangement for epiretinal prostheses arXiv:2203.02493

[24] Jepson L H *et al* 2014 Spatially patterned electrical stimulation to enhance resolution of retinal prostheses *J. Neurosci.* **34** 4871–81

[25] Grosberg L E *et al* 2017 Activation of ganglion cells and axon bundles using epiretinal electrical stimulation *Journal of Neurophysiology* **118** 1457–71

[26] Tandon P *et al* 2021 Automatic Identification of Axon Bundle Activation for Epiretinal Prosthesis *IEEE Trans. Neural Syst. Rehabil. Eng.* **29** 2496–502

[27] Vilkhu R S *et al* 2021 Spatially patterned bi-electrode epiretinal stimulation for axon avoidance at cellular resolution *J. Neural Eng.* **18** 066007

[28] Tong W *et al* 2019 Improved visual acuity using a retinal implant and an optimized stimulation strategy *J. Neural Eng.* **17** 016018

[29] Gonzalez-Calle A and Weiland J D 2016 *Evaluation of effects of electrical stimulation in the retina with Optical Coherence Tomography* **2016** 6182–5

[30] Chang Y C, Ghaffari D H, Chow R H and Weiland J D 2019 Stimulation strategies for selective activation of retinal ganglion cell soma and threshold reduction *J. Neural Eng.* **16** 026017

[31] Ghaffari D H *et al* 2020 The effect of waveform asymmetry on perception with epiretinal prostheses *J. Neural Eng.* **17** 045009

[32] Hertzmann A 2003 A survey of stroke-based rendering *IEEE Computer Graphics and Applications* **23** 70–81

[33] Mellor J F J *et al* 2019 *Unsupervised Doodling and Painting with Improved SPIRAL* arXiv:1910.01007

[34] Zheng N, Jiang Y and Huang D 2019 StrokeNet: A Neural Painting Environment *Int. Conf. on Learning Representations*

[35] Zhou T *et al* 2018 Learning to doodle with stroke demonstrations and deep Q-Networks *British Machine Vision Conf.*

[36] Ha D and Eck D 2017 *A Neural Representation of Sketch Drawings* arXiv:1704.03477

[37] Huang Z, Heng W and Zhou S 2019 *Learning to Paint With Model-based Deep Reinforcement Learning* arXiv:1903.04411

[38] Becker M, Braun M and Eckmiller R 1998 Retina implant adjustment with reinforcement learning *Proc. of the 1998 IEEE Int. Conf. on Acoustics, Speech and Signal Processing, ICASSP'98 (Cat. No. 98CH36181)* (Seattle, WA: IEEE) 2, 1181–4

[39] Eckmiller R, Neumann D and Baruth O 2005 Tunable retina encoders for retina implants: why and how *J. Neural Eng.* **2** S91–104

[40] Küçükoglu B, Rueckauer B, Ahmad N, de Ruyter van Steveninck J, Güçlü U and van Gerven M 2022 Optimization of neuroprosthetic vision via end-to-end deep reinforcement learning *International Journal of Neural Systems* **32** 2250052

[41] Becker M and Eckmiller R 1997 Spatio-temporal filter adjustment from evaluative feedback for a retina implant *Int. Conf. on Artificial Neural Networks* 1181–6

[42] Barto A G and Dietterich T G 2004 Reinforcement learning and its relationship to supervised learning *Handbook of Learning and Approximate Dynamic Programming* **10** 45–63

[43] Ha D and Schmidhuber J 2018 *Recurrent World Models Facilitate Policy Evolution* arXiv:1809.01999

[44] Sutton R S and Barto A G 2018 *Reinforcement Learning: An Introduction* (The MIT Press)

[45] Brockman G *et al* 2016 *OpenAI Gym* arXiv:1606.01540

[46] Jansonius N M *et al* 2009 A mathematical description of nerve fiber bundle trajectories and their variability in the human retina *Vis. Res.* **49** 2157–63

[47] Deng L 2012 The mnist database of handwritten digit images for machine learning research *IEEE Signal Process Mag.* **29** 141–2

[48] Snellen H 1862 *Probebuchstaben zur bestimmung der sehscharfe* (Van de Weijer)

[49] Haarnoja T, Zhou A, Abbeel P and Levine S 2018 *Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor* arXiv:1801.01290

[50] Silver D, Lever G, Heess N, Degris T, Wierstra D and Riedmiller M 2014 Deterministic policy gradient algorithms *Int. conf. on machine learning. PMLR* 387–95

[51] Lillicrap T P *et al* 2019 *Continuous control with deep reinforcement learning* arXiv:1509.02971

[52] Liu R *et al* 2018 An intriguing failing of convolutional neural networks and the coordconv solution *Advances in neural information processing systems* 31

[53] Liang E *et al* 2018 RLlib: Abstractions for distributed reinforcement learning *Int. Conf. on Machine Learning. PMLR* 3053–62

[54] Mnih V *et al* 2015 Human-level control through deep reinforcement learning *Nature* **518** 529–33

[55] Kingma D P and Ba J 2014 Adam: a method for stochastic optimization arXiv:1412.6980

[56] Arjovsky M, Chintala S and Bottou L 2017 Wasserstein generative adversarial networks *Proc. of the 34th Int. Conf. on* .

vol. 70 of Proceedings of Research. PMLR ed D Precup and Y W Teh 214–23

[57] Cuturi M 2013 *Sinkhorn Distances: Lightspeed Computation of Optimal Transport Advances in neural information processing systems* 26

[58] Bruna J, Sprechmann P and LeCun Y 2015 *Super-resolution with deep convolutional sufficient statistics* arXiv:1511.05666

[59] Ledig C *et al* 2017 Photo-realistic single image super-resolution using a generative adversarial network *Proc. of the IEEE conf. on computer vision and pattern recognition* 4681–90

[60] Zhao H, Gallo O, Frosio I and Kautz J 2015 *Loss Functions for Neural Networks for Image Processing* arXiv:1511.08861

[61] Wang Z, Bovik A C, Sheikh H R and Simoncelli E P 2004 Image quality assessment: from error visibility to structural similarity *IEEE Trans. Image Process.* **13** 600–12

[62] Palanker D, Huie P, Vankov A, Aramant R, Seiler M, Fishman H *et al* 2004 Migration of retinal cells through a perforated membrane: implications for a high-resolution prosthesis *Investigative Ophthalmol. Vis. Sci.* **45** 3266–70

[63] Metz L, Poole B, Pfau D and Sohl-Dickstein J 2016 *Unrolled generative adversarial networks* arXiv:1611.02163

[64] Arjovsky M, Chintala S, Bottou L and Wasserstein G A N 2017 arXiv:1701.07875

[65] Nanduri D, Fine I, Horsager A, Boynton G M, Humayun M S, Greenberg R J *et al* 2012 Frequency and amplitude modulation have different effects on the percepts elicited by retinal stimulation *Investigative Ophthalmol. Vis. Sci.* **53** 205–214

[66] Relic L, Zhang B, Tuan Y L and Beyeler M 2022 Deep learning-based perceptual stimulus encoder for bionic vision *Assoc. for Computing Machinery: Augmented Humans 2022* 323–5