2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 978-1-7281-1179-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/EMBC46164.2021.9629869

Using Biologically-inspired Image Features to Model Retinal Response: Evidence from Biological Datasets

Nikos Melanitis*, Giorgos Nakopoulos‡, Antonio Lozano†, Cristina Soto-Sanchez†, Eduardo Fernandez† and Konstantina S. Nikita‡

Abstract— Retinal models are needed to simulate the translation of visual percepts to Retinal Ganglion Cells (RGCs) neural spike trains, through which visual information is transmitted to the brain. Restoring vision through neural prostheses motivates the development of accurate retinal models. We integrate biologically-inspired image features to RGC models. We trained Linear-Nonlinear models using response data from biological retinae. We show that augmenting raw image input with retina-inspired image features leads to performance improvements: in a smaller (30sec. of retina recordings) set integration of features leads to improved models in approximately $\frac{2}{3}$ of the modeled RGCS; in a larger (4min. recording) we show that utilizing Spike Triggered Average analysis to localize RGCs in input images and extract features in a cell-based manner leads to improved models in all (except two) of the modeled RGCs.

Index Terms— retinal model, bio-inspired features, computer vision, retinal prosthesis, biological data

I. Introduction

Retina processes the information in a complex way before relaying it, through numerous Retinal Ganglion Cell (RGC) types, to the brain [1]. Models of the retina have received much attention, especially in light of neural-prosthetic approaches aiming to restore vision [2], [3]-predicting the retina response is essential to properly encode a visual scene to electrical neural stimulations delivered by neural prostheses.

Simple Linear-Nonlinear (LN) models predict accurately retina's response to white noise [4], more recent deep learning models accurately reproduce retina's response to natural images [5], [6] while bio-inspired image features, based on RGC functions, have been integrated to retina models of both linear and nonlinear RGC types [7]. Still, validation of bio-inspired image features has been limited to artificial data [7]. Integration of features to more RGC models needs to be further explored.

In this study we validate using bio-inspired image features to improve retina models. We use biological data recordings from RGCs, as a more solid validation procedure. In order to do so, we modify feature extraction to account for unknown biological RGC response parameters and show how computational neuroscience methods, proper experimental stimulus design and AI modeling methods may be utilized to properly extract the image features. We use LN models which can be trained easily with a moderate amount of data and are straightforward to use with features as input. We show that combining image input with the extracted features improves RGC models and highlight that determining RGC spatial and temporal response characteristics is critical to properly extract image features.

II. Methods

A. Datasets

We used two datasets of RGC responses to natural images:(i) a Salamander[1] set, (ii) a Mouse[2] set.

Salamander set contains recordings from 21 biological RGCs. Natural Images of 50×50 pixels are projected on the retina for 1 s. each and for a total duration of 30 s. Eye-jitter motion was added in image projection. To reduce noise in the recordings, RGC response was averaged over 112 repeated projections of the image sequence on the retina. In total, we have 3003 recordings per cell.

Mouse set contains recordings from 60 biological RGCs. Natural Images of 50×50 pixels are projected on the retina for 50 ms each and for a total duration of 4 min. The sequence starts with 300 ms of darkness. The set contains static images only. We record cell response every 10 ms. In total, we have 24479 recordings per cell.

Reliably-recorded RGCs: In 'Mouse' set, we get twelve reliably-recorded RGCs, through Spike Triggered Average (STA) analysis [8]. Errors in raw data processing (e.g., in spike sorting) and/or at the retina preparation may corrupt the biological recordings. We trained a Convolutional Neural Network (CNN) model [5] on 'Mouse' set and then fed white noise sequences to the model to get an unbiased Receptive Field (RF) estimate through STA [8]. Reliable cells were selected based on spatial (center-surround antagonism) and temporal (biphasic response) STA characteristics. The STA properties of RGCs have been documented in the literature [8].

*N. Melanitis is with the School of Electrical and Computer Engineering, National Technical University of Athens, Greece nmelan@biosim.ntua.gr

† E.Fernandez, C. Soto-Sanchez and A. Lozano are with Instituto de Bioingenieria, Universidad Miguel Hernandez, Alicante, Spain

‡ K.S. Nikita and G. Nakopoulos are with the School of Electrical and Computer Engineering, National Technical University of Athens, Greece

[1]Available at https://github.com/lmcintosh/deep-retina-tutorial

[2]Biological recordings at Professor's E. Fernandez lab

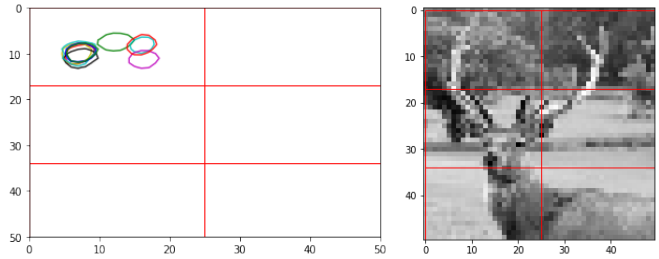


Fig. 1: Retinal Ganglion Cells (RGCs) localization and RGC-parametrized Feature extraction: We show the Receptive Fields (RFs) of reliably-recorded RGCs in 50×50 pixels Images. RGCs are contained in a 17×25 pixels Image sub-region (left). In RGC-parametrized feature extraction, we feed the top-left image sub-region to the LN model, which contains all the RFs (right). RGCs localization was conducted by STA analysis of RGC response to white noise, using a CNN model of the RGCs

B. Feature Extraction

We trained models with input: (i) unprocessed images ('I'), (ii) Features ('F') and (iii) Features and Images ('FI'). Feature extraction (FE) follows our previous work [7]. Since spatiotemporal RGC properties are not available in the datasets, we set an RF size of 18×18 pixels [7], tile the images with non-overlapping RFs that cover them entirely and extract features in all the RFs defined. The aforementioned modification increases the total number of extracted features in each image, which we limit by dropping all the 'Detector' features [7] and extracting: Difference of Gaussians, Canny edges, optical flow and uniformity features. This way, we extract 7527 features in 50×50 pixels images. This is our baseline, 'Full Image' FE approach.

FE in Still Images: We modified FE in the static-images 'Mouse' set, to remove motion-based features (Optic Flow, Temporal Uniformity). This way, we extract 2059 features in 50×50 pixels images.

RGC-parametrized FE: We localized RGC positions in Images of the 'Mouse' set. We determined the RF spatial positions of the twelve reliable RGCs (see 'Reliably-recorded RGCs' above and fig. 1) and isolated a $17 \times$ 25pixels image region that contained all the RFs. We extracted 1281 features (429 in the 'Still Images' case) in the aforementioned image sub-region.

C. Model description and training

To predict RGCs firing rates, we trained Linear-Nonlinear-Poisson (LN) models with a parametric 'Softplus' nonlinearity optimizing the Poisson loss function [4], specifically the mean value of the error:

$$\hat{y} - y \cdot \log \hat{y} \tag{1}$$

where $\hat{y}$ is the estimate of y. LN models generate RGC neural spikes from an inhomogeneous Poisson process with rate λ:

$$\lambda(t) = f(\vec{k} \cdot \vec{x}(t)) \tag{2}$$

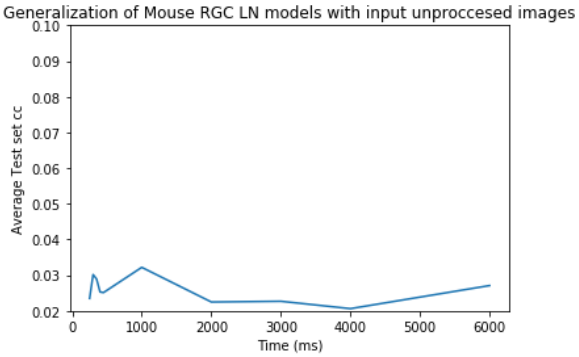
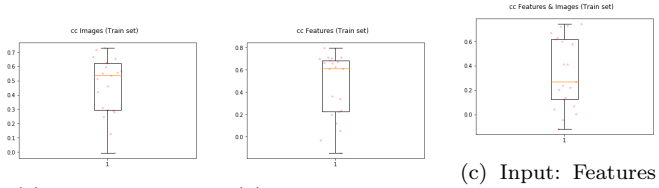


Fig. 2: Mean Testing set performance (correlation coefficient) of reliably-recorded RGCs LN models with Image input over varying temporal filter extent (t_f) values



(a) Input: Images (b) Input: Features (c) Input: Features and Images

Fig. 3: Performance (cc) of LN models in predicting RGC response. Points in the graph denote the cc of specific RGCs. The mean value and the range of three standard deviations are plotted. Results for 'Salamander' RGCs, Training set

where f is a nonlinearity and $\vec{k}$ a linear filter acting on input $\vec{x}$.

We applied the numerical stochastic gradient descent optimization algorithm and did early-stopping in training. We did a $70\% - 30\%$ Training-Testing set split.

We evaluate our models with the correlation coefficient (cc) that has been used extensively in the literature [5], [6], [9]:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{3}$$

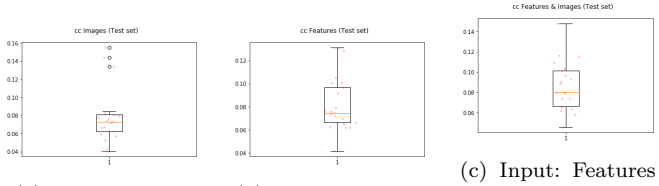
where n is the sample size, x_i, y_i are sample points and $\bar{x}, \bar{y}$ the respective mean values.

In LN models, we need to set the temporal extent t_f of the linear filter: to what temporal extent past inputs influence RGC response? In 'Salamander' set the scientists that gathered the data determined t_f value (400 ms). In 'Mouse' set we examined cc of LN models with 'I' input and varying t_f (fig. 2, observing only small variations in model performance. We consequently set $t_f = 400$ ms, the value at which our CNN models (see 'Reliably-recorded RGCs') yielded the anticipated (center-surround spatial and biphasic temporal) STA filters shape.

III. Results and Discussion

A. LN models with features show improved performance in a limited-size, natural-image dataset

In figs. 3, 4, 5 we show our results on 'Salamander' dataset. We applied early-stopping in model training.



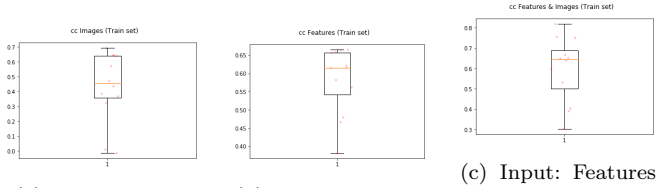
(a) Input: Images (b) Input: Features (c) Input: Features and Images

Fig. 4: Performance (cc) of LN models in predicting RGC response. Points in the graph denote the cc of specific RGCs. The mean value and the range of three standard deviations are plotted. Results for 'Salamander' RGCs, Testing set

(a) Image Input vs Features Input

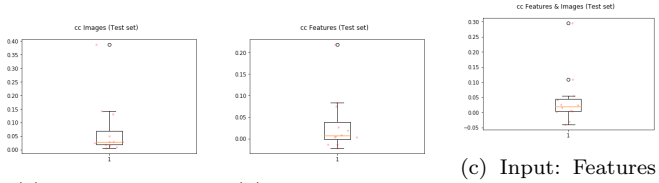
(b) Image Input vs Features-and-Image Input

Fig. 5: Performance (cc) of LN models. We compare models of each RGC over different input representations-Images, Features, Features and Images



(a) Input: Images (b) Input: Features (c) Input: Features and Images

Fig. 6: Performance (cc) of LN models in predicting RGC response. Points in the graph denote the cc of specific RGCs. The mean value and the range of three standard deviations are plotted. Results for 'Mouse' RGCs, Training set



(a) Input: Images (b) Input: Features (c) Input: Features and Images

Fig. 7: Performance (cc) of LN models in predicting RGC response. Points in the graph denote the cc of specific RGCs. The mean value and the range of three standard deviations are plotted. Results for 'Mouse' RGCs, Testing set

In figs. 3, 4 we observe that all our models-ie with 'I', 'F' and 'FI' input, performed according to our expectations (cc ≈ 0.10) for the dataset size [5].

We compare models on an RGC-by-RGC basis in fig. 5. Models with features input ('F', 'FI') are improved in comparison to the model with image input ('I'), for each cell (point in fig. 5) with a negative margin. 'F' models overperform 'I' models in 14 of the total 21 RGCs.

Still, not all 'FI' models overperform 'I' models, as expected by observing that 'FI' input is 'I' input augmented by 'F'. We conclude that a larger dataset is needed to fit a 'FI' model, which is more complex (i.e. has more parameters) than an 'I' model.

B. Cell- and dataset- specific modifications in feature extraction lead to improved LN models

In Table I we present the Testing and Training set performance on 'Mouse' dataset for three -'Full Image', 'RGC-parametrized', 'Still Images'- feature extraction approaches.

We observe that performance of the baseline ('Full Image') approach is below our expectations [5], even for LN models with Image ('I') Input. The lower performance of 'I' models, compared to 'Salamander' set, directs us (i) to investigate dataset differences and to examine how we may address them, or (ii) to point to biological differences between the RGCs, which make unfavorable the application of our approach in models of the mouse retina. We see that noise suppression (by repeating input images) in the recordings and eye-jitter motion favoured FE in 'Salamander' set.

	Full Image			RGC-parametrized			Still Images	
	I	F	FI	I	F	FI	F	FI
Train	0.30	0.20	0.14	0.43	0.48	0.39	0.49	0.45
Test	0.03	0.03	0.03	0.07	0.03	0.03	0.06	0.09

TABLE I: Performance (cc) of LN models in predicting RGC response. Three feature extraction approaches, the baseline 'Full Image' approach, RGC-parametrized extraction ('RGC-parametrized') and removing motion-based features ('Still Images'), are compared. Results for 'Mouse' RGCs

We start by applying RGC-parametrized feature extraction [7]. All twelve RGCs in the 'Mouse' set are concentrated in an image region (fig. 1), which we isolate and use as model and feature extraction input. Doing so, we guide the LN filters to the image region that the modeled RGCs 'sense', and also reduce the number of parameters (filter weights) in our models. In previous work, we have applied RGC-parametrized feature extraction in a spatially more restrictive way, isolating the image region that spans a single RF, which further lowers the dimensionality of extracted features to around twenty [7]. We see that informed-ie Cell Based- FE is important to integrate features to improved RGC models. In Table I, we see an improvement in 'I' models, but feature-based approaches ('F', 'FI') still under-perform.

We next turn to the extracted features, remove all motion-based features and maintain the RGC-parametrized approach. While Salamander set has a 'jitter' motion, Mouse set has no motion at all. We observe (Table I) that feature approaches are now improved. We note that 'FI' approach now dominates 'I' approach, with only two cases of 'I' models performing better than 'FI' models (fig. 8). However, 'F' models show a damped improvement-low margins (fig. 8) that

(a) Image Input vs Features Input

(b) Image Input vs Features-and-Image Input

Fig. 8: Performance (cc) of LN models. We compare models of each RGC over different input representations- Images, Features, Features and Images

are very concentrated around 0 -in 'Mouse' set compared to 'Salamander' set- in which we see larger margins (fig. 5). This highlights the limitations of applying feature extraction in still Images datasets.

Stimulus design- i.e. selecting appropriate stimuli to project on the retina and make recordings- can facilitate feature extraction. White noise stimuli simplify t_f identification (see fig. 2, a sweep of t_f values provided limited information) and RGC localization through an unbiased STA estimate of RGC response. More features can be extracted when motion is present in the stimuli.

Overall, we see that 'FI' approach was more suitable for the 'Mouse' set, which is larger and so can be used to train a complex (high number of parameters) model but lacks any motion, while 'F' approach excelled in the smaller 'Salamander' set in which we were able to extract all (motion and non-motion dependent) features.

IV. Conclusions and Future Work

We have integrated biologically-inspired image features in retina models using biological recordings to train and evaluate our approach. We show that model performance improves when features are used as model input. We argue that characterizing RGCs spatiotemporal response properties is a crucial step: localizing RGCs improves feature extraction, determining the extent of temporal response is important in models of RGCs. Dataset characteristics, as the inclusion of movement, determine which image features we can extract and affect the impact on performance of extracting image features. Consequently, the importance of a close collaboration between experimentalists and engineers is highlighted by this study.

In future work, we should extend our approach from simple (LN) models to complex (Deep learning-DL) models which can reliably reproduce retina's response to natural images[6]. Feature fusion approaches [10], [11], [12] should be considered to explore how image features are integrated in DL models. Another important direction entails the explainability of our models, revealing which features and/or image regions influence the retina's output, at varying stimulus conditions [13]. Towards improved visual prostheses, accurate retina models should be combined with smart image processing algorithms to implement important visual functions, as for example methods to predict where visual attention is directed [14].

References

[1] B. Roska and M. Meister, "The retina dissects the visual scene into distinct features," The new visual neurosciences, vol. 843, pp. 163–182, 2014.

[2] S. Niketeghad and N. Pouratian, "Brain machine interfaces for vision restoration: the current state of cortical visual prosthetics," Neurotherapeutics, vol. 16, no. 1, pp. 134–143, 2019.

[3] E. Fernandez, "Development of visual neuroprostheses: trends and challenges," Bioelectronic medicine, vol. 4, no. 1, pp. 1–8, 2018.

[4] J. W. Pillow, J. Shlens, L. Paninski, A. Sher, A. M. Litke, E. Chichilnisky, and E. P. Simoncelli, "Spatio-temporal correlations and visual signalling in a complete neuronal population," Nature, vol. 454, no. 7207, pp. 995–999, 2008.

[5] L. McIntosh, N. Maheswaranathan, A. Nayebi, S. Ganguli, and S. Baccus, "Deep learning models of the retinal response to natural scenes," in Advances in neural information processing systems, pp. 1369–1377, 2016.

[6] A. Lozano, C. Soto-Sanchez, J. Garrigos, J. J. Martínez, J. M. Ferrández, and E. Fernandez, "A 3d convolutional neural network to model retinal ganglion cell's responses to light patterns in mice," International journal of neural systems, vol. 28, no. 10, p. 1850043, 2018.

[7] N. Melanitis and K. S. Nikita, "Biologically-inspired image processing in computational retina models," Computers in biology and medicine, vol. 113, p. 103399, 2019.

[8] E. Chichilnisky, "A simple white noise analysis of neuronal light responses," Network: Computation in Neural Systems, vol. 12, no. 2, pp. 199–213, 2001.

[9] A. Ecker, F. Sinz, E. Froudarakis, P. Fahey, S. Cadena, E. Walker, E. Cobos, J. Reimer, A. Tolias, and M. Bethge, "A rotation-equivariant convolutional neural network model of primary visual cortex," in Seventh International Conference on Learning Representations (ICLR 2019), pp. 1–11, 2019.

[10] Y. Wang, B. Song, P. Zhang, N. Xin, and G. Cao, "A fast feature fusion algorithm in image classification for cyber physical systems," IEEE Access, vol. 5, pp. 9089–9098, 2017.

[11] S. Hosseini, S. H. Lee, and N. I. Cho, "Feeding hand-crafted features for enhancing the performance of convolutional neural networks," arXiv preprint arXiv:1801.07848, 2018.

[12] J. Dapello, T. Marques, M. Schrimpf, F. Geiger, D. D. Cox, and J. J. DiCarlo, "Simulating a primary visual cortex at the front of cnns improves robustness to image perturbations," BioRxiv, 2020.

[13] M. Athanasiou, K. Sfrintzeri, K. Zarkogianni, A. C. Thanopoulou, and K. S. Nikita, "An explainable xgboost–based approach towards assessing the risk of cardiovascular disease in patients with type 2 diabetes mellitus," in 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE), pp. 859–864, 2020.

[14] A. Alevizaki, N. Melanitis, and K. Nikita, "Predicting eye fixations using computer vision techniques," in 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE), pp. 309–315, 2019.