GROUP-PATCH BASED CLASSIFICATION AND ASYMPTOTIC PREDICTING IMBALANCED NEURON SPIKES

Mingli Zhang[1,3] *Dongsheng Xiao*[2] *Timothy H. Murphy*[2] *Jean-Baptiste Poline*[1,4⋆] *Alan Evans*[1,⋆]

[1]Montreal Neurological Institute, Mcgill University, Montreal, Canada
[2] Djavad Mowafaghian Centre for Brain Health, University of British Columbia, Vancover, Canada
[3] Shandong Co-innovation Center of Future Intelligent Computing, Yantai, China
[4] Helen Wills Neuroscience Institute, University of California, Berkeley, California, USA
⋆ Co-last author

Abstract – **The cerebral cortex is connected to subcortical structures through multiple hierarchically organized descending and ascending pathways. Cortical representations of subcortical neural activity reflect embedded complex spatiotemporal dynamics. Past studies extensively examined the molecular, cellular and circuit properties of intracortical and cortico-subcortical pathways. It remains poorly understood how the spiking activity of a cortical or subcortical neuron is associated with cortex-wide network dynamics. In this study, we use simultaneously recorded mesoscale calcium imaging and behavior video as multi-modality for predicting neuron spiking on awake mice. A novel group patch wise classification and asymptotic model is proposed to address the ultra imbalanced spikes prediction problem. Experiments demonstrate the efficacy of the proposed classification and asymptotic predicting strategies. Our empirical results reveal that the spike activities of neurons were associated with distinct cortical calcium signals and behavior information,which was observed that ongoing neural cortical activity encodes a high-dimensional latent signal of behavior and sensory.**

I. INTRODUCTION

One of the important topics in neuroscience is designing effective encoding model and applying to neural spiking prediction [1], [2]. A shared variance component analysis method was applied for the estimation of neural population's variance reliably encod- ing a latent signal, alternatively, ongoing activity could be related to behavioral and cognitive states [2]. Coordinated spiking activity in local networks and across multiple cortical areas is considered to provide the neural basis for understanding cognition and adaptive behavior with point process history models (Bayesian estimation) in [3].

Single neuron spiking activity is associated with intrinsic biophysical diversification and the interaction among neurons in neuronal ensembles. Sensory and motor processing engages extended intracortical and cortico-subcortical networks carrying feedforward and feedback information, including intracortical connections between a variety of cortical modules [4]. How the cerebral cortex is connected with subcortical structures through multiple hierarchically organized descending and ascending pathways. Cortex-wide representations of cortical and subcortical neural activity thus reflect embedded complex spatiotemporal dynamics. It remains poorly understanding on how spikes activity of a cortical or subcortical neuron is associated with cortex-wide network dynamics. Recent advances in large-scale optical imaging methods [5] provide opportunities to reveal the cortical ensemble representation for neuronal activity. In this study, we will use multimodal data that simultaneously recorded wide-field calcium imaging, behavior video and multi-unit electrophysiology (in the barrel cortex, thalamus, hippocampus, striatum, and the ventral tegmental area) in awake and anesthetized mice for single neuron spiking prediction, as shown in Fig 1.

Machine learning and deep learning achieved considerable achivements in decoding neural representations [6]. Such as the deep convolutional neural networks (CNNs) have been successfully applied to infer stimulus selectivity from inferotemporal and prefrontal spikes [7]. In this paper, we would propose an asymptotic classification approach to learn the low-dimensional cortical calcium signals and behavior videos to predict the single neuronal spike activity.

To accurately predict single neuronal with multimodality is challenging topic in the aspect that 1) appropriately fuse the multimodal behavior video and cortical calcium image dataset 2) the distribution of each spike class is an ultra imbalance, more than 80 percent of spikes are motionless signals(zeros), as shown in Fig 2 a). 3) The behavior video and cortical calcium image are in time-sequence format, as shown in Fig 2 b), each spike is not only related to the corresponding behavior and cortical calcium image, but also their neighbors, what's more, arranging the video behavior features and cortical calcium images corresponding to neural spikes,may have misplacement issue. To address the aforementioned challenging, in this paper, we proposed a group-patch based classification and asymptotic predicting

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

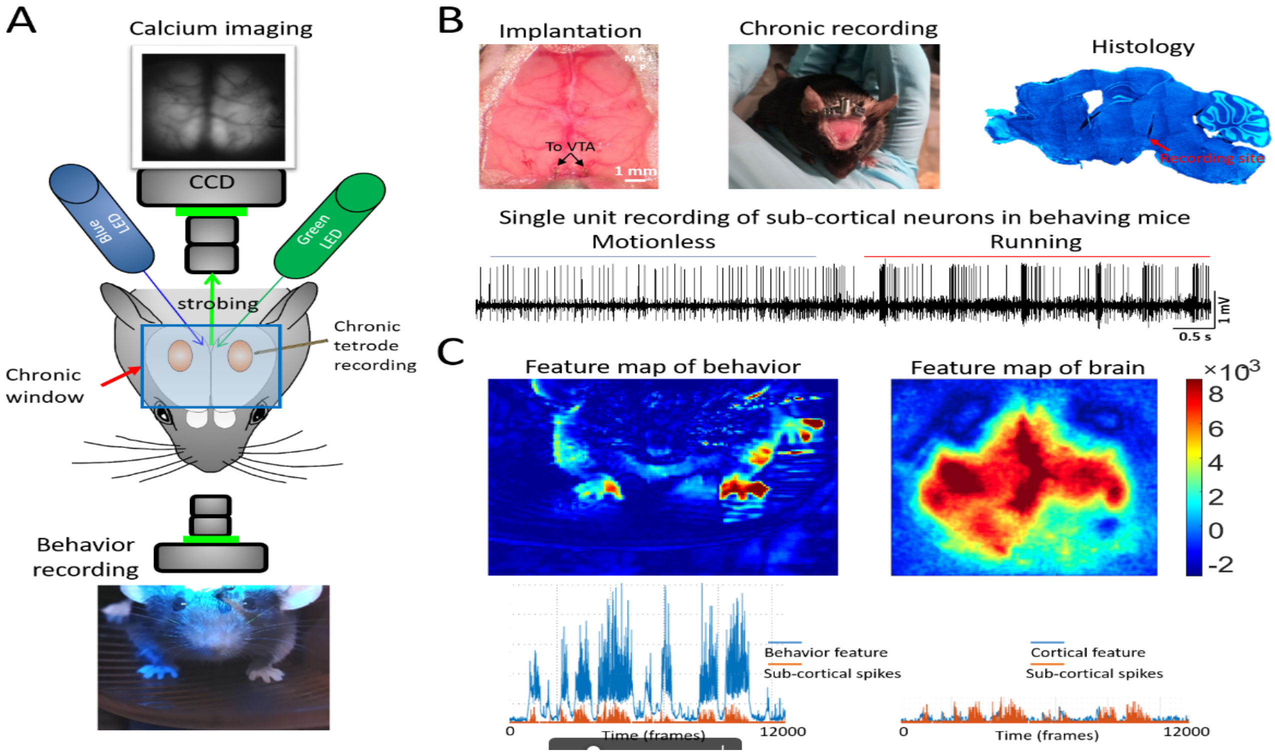

Fig. 1: Calcium imaging, behavior video and single neuron spiking recording processing

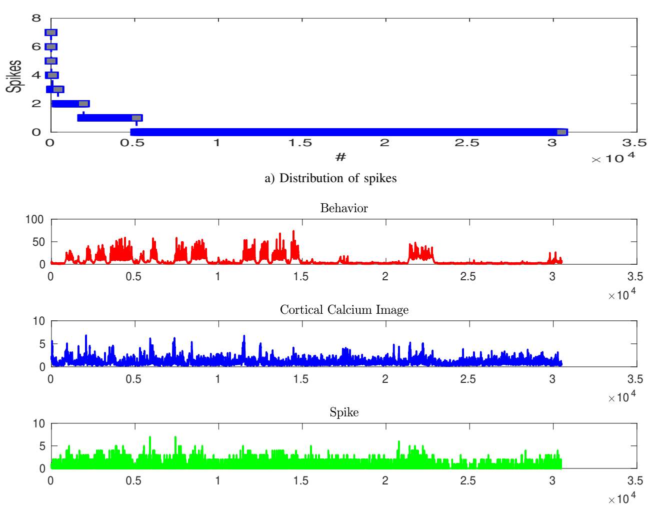

Fig. 2: The distribution of Spikes and principal components of Image and Behavior. The correlation of spikes and principal components of image is $r_{si} = 0.0979$, the correlation of spikes and principal components of behavior is $r_{sb} = 0.2478$.

approach. The main contributions of this work are listed as follows.

- **Novel method:** Motivated by non-negative projective dictionary learning (NPDL) algorithms of Zhang *et al.* [8], we propose the tightest robust principal component analysis (rPCA) based batch setting estimation model to address the aforementioned multi-modal fusion and single neuronal spike prediction problem. Group-patch approaches have been shown to be useful in numerous priors learning problems like image processing [9]. While NPDL applies instance with fully supervised learning, to consider the local similarity of instance (feature) and consistency of spike and features, we propose a feature overlap local group patch approach. In addition to improve the prediction accuracy of minority spikes on the ultra imbalanced dataset, we propose a group patch general classification guided asymptotic regression approach to predict the minority spikes.
- **Application:** We evaluate our group patch rPCA based subspace estimation framework on the prediction of single neural spikes, using recorded wide-field calcium imaging, behavior video [5], the results show a higher accuracy compared to NDPL and the most commonly used classical model. Our experiments were implemented by features of calcium imaging and behavior video, which are important for predicting single neuronal spikes, the impact of each modality on the prediction were also studied. The association between the predicted single neuronal spikes and calcium imaging is also investigated with the proposed model.

II. PROBLEM FORMULATION

In this section, we formulate the single neural spikes classification problem with a class-wise patch based subspace estimation approach, applying group patch based robust principal component analysis (rPCA) in batch setting. To make full use of the local similarity of the instance (feature) and consistency of the neighbor spikes and features, we propose an overlap local feature patch approach. The rPCA classification guided asymptotic regression approach is proposed to predict the minority spikes on the ultra imbalanced multimodality dataset.

II-A. rPCA with Batch Setting

The rPCA based subspace estimation problem is commonly used in signal processing and statistical learning, with the hypothesis that data are with zero-mean, and we estimate the principal subspace of the covariance matrix from the data. Compared with dictionary learning approach, the proposed rPCA based approach aims to learn the coding coefficients and dictionary of each class separately, we incorporate a structured synthesis dictionary $\mathbf{D}$ for reconstructing the data, and gain a projection $\mathbf{P}$ for extracting the discriminative codes from data.

We treat the modeling of single neural spikes as a group patch based robust principal component analysis (rPCA) learning problem over K classes, and for the major spikes, each class is a spike group, while all the minority spikes are assigned to the final group K. Such as the data used in this work $C = [25390\ 3168\ 1523\ 33969\ 15\ 1\ 2]$, we assign C_1 and C_2 into class $k = 1$ and $k = 2$, respectively, the left are assigned into one final class $K = 3$. Let $\mathbf{X} = [\mathbf{X}_1, \cdots, \mathbf{X}_k, \cdots, \mathbf{X}_K]$ be one modality samples from all classes and $\mathbf{X}_k \in \mathbb{R}^{S\times N_k}$ is the data samples of the k-th class, where, S is the vector format features for one spike, it is re-arranged the several local patches of the spike, N_k is the number of spikes in the k group. Following the former works [8], [10], for each class k, we introduce an analysis dictionary $\mathbf{P}_k \in \mathbb{R}^{M\times S}$ to project the data into a coefficient matrix and a synthesis dictionary $\mathbf{D}_k \in \mathbb{R}^{S\times M}$ to reconstruct the data, where, M is the number of spikes in the class and S is the dimension of each feature, such that for all classes we have $\mathbf{P} = [\mathbf{P}_1, \cdots, \mathbf{P}_k, \cdots, \mathbf{P}_K]$ and $\mathbf{D} = [\mathbf{D}_1, \cdots, \mathbf{D}_k, \cdots, \mathbf{D}_K]$. With these dictionaries, one can perform data modeling with rPCA:

$$\begin{aligned} \underset{\mathbf{D}_k,\mathbf{P}_k=\mathbf{D}_k^\top}{\arg\min} \sum_{k=1}^{K} & \|\mathbf{X}_k - \mathbf{D}_k\mathbf{P}_k\mathbf{X}_k\|_F^2 + \lambda_1\|\mathbf{P}_k\mathbf{X}_k\|_{w,*} \\ & + \sum_{\overline{k}\in r(k)} \left(\lambda\|\mathbf{P}_k\mathbf{X}_{\overline{k}}\|_F^2 + \lambda_2\|\mathbf{P}_k\mathbf{X}_{\overline{k}}\|_1\right) \\ \text{s.t.} \quad & \mathbf{D}_k^\top\mathbf{D}_k = \mathbf{I}, \quad k = 1, ..., K. \end{aligned} \tag{1}$$

where $\|\cdot\|_F$ is the Frobenius norm, $\overline{\mathbf{X}}_k = [\mathbf{X}_1, \cdots, \mathbf{X}_{k-1}, \mathbf{X}_{k+1}, \cdots, \mathbf{X}_K]$, and $\lambda, \lambda_1, \lambda_2 > 0$ control the trade-off between the reconstruction accuracy and regularization terms. $\|\mathbf{P}_k\overline{\mathbf{X}}_k\|_F^2$ is used to forcing $\mathbf{P}_k\mathbf{X}_{\overline{k}}$ towards small or zero values for any other class $\overline{k} \in \{\overline{k} : |k - \overline{k}| \neq 0\}$. In this model, $\mathbf{P}_k$ projects the samples $\mathbf{X}_k$ into an encoding coefficient matrix $\mathbf{A}_k = \mathbf{P}_k\mathbf{X}_k$. One important issue for PCA in batch setting is deteministic, which can be tedious. Hence we propose to apply a low-rank regularization over the coefficient matrix with l_1 norm. $\mathbf{A}_{\overline{k}} = \mathbf{P}_k\mathbf{X}_{\overline{k}}$ for $\overline{k} \in r(k)$, which transforms the equation (1) into:

$$\begin{aligned} \underset{\mathbf{D}_k,\{\mathbf{A}_k,\{\mathbf{A}_{\overline{k}}:\overline{k}\in r(k)\}\}}{\arg\min} \sum_{k=1}^{K} & \|\mathbf{X}_k - \mathbf{D}_k\mathbf{D}_k^\top\mathbf{X}_k\|_F^2 + \lambda_1\|\mathbf{A}_k\|_{w,*} \\ & + \sum_{\overline{k}\in r(k)} \left(\lambda\|\mathbf{P}_k\mathbf{X}_{\overline{k}}\|_F^2 + \lambda_2\|\mathbf{A}_{\overline{k}}\|_1\right) \\ \text{s.t.} \quad & \mathbf{D}_k^\top\mathbf{D}_k = \mathbf{I},\ \mathbf{A}_k = \mathbf{P}_k\mathbf{X}_k, \\ & \mathbf{A}_{\overline{k}} = \mathbf{P}_k\mathbf{X}_{\overline{k}},\ \overline{k} \in r(k)\ k = 1, ..., K. \end{aligned} \tag{2}$$

To learn sets $\mathbf{D}_k$, $\mathbf{A}_k$ and $\mathbf{A}_{\overline{k}}$, we present an efficient optimized approach in the following Training section.

II-B. Training

We introduce the auxiliary variable matrices, $\{\mathbf{Z}_{1,k}\}$ and $\{\mathbf{Z}_{2,k}\}$, as equality constraints into the model 2 for optimizing $\mathbf{D}_k$,$\mathbf{A}_k$ and $\mathbf{A}_{\overline{k}}$ iteratively.

Updating D: Given fixed $\{\mathbf{A}_k, \{\mathbf{A}_{\overline{k}} : \overline{k} \in r(k)\}, \mathbf{Z}$, the minimization over each $\mathbf{D}_k$ is as follows:

$$\underset{\mathbf{D}_k}{\arg\min}\ \|\mathbf{X}_k - \mathbf{D}_k\mathbf{D}_k^\top\mathbf{X}_k\|_F^2,\ \text{s.t. } \mathbf{D}_k^\top\mathbf{D}_k = \mathbf{I}$$

which can be rewritten as follows:

$$\underset{\mathbf{D}_k}{\arg\max}\ \text{tr}(\mathbf{D}_k^\top\mathbf{X}_k\mathbf{X}_k^\top\mathbf{D}_k),\ \text{s.t. } \mathbf{D}_k^\top\mathbf{D}_k = \mathbf{I}, \tag{3}$$

It can be solved with eigenvalue decomposition of $\mathbf{X}_k\mathbf{X}_k^\top$, with $\mathbf{X}_k\mathbf{X}_k^\top = \hat{\mathbf{D}}_k\Sigma\hat{\mathbf{D}}_k^\top$ [11].

Updating A: Given fixed $\{\mathbf{D}, \{\mathbf{A}_{\overline{k}} : \overline{k} \in r(k)\}, \mathbf{Z}\}$, $\mathbf{A}_k$ is solved as follows:

$$\underset{\mathbf{A}_k}{\arg\min}\ \|\mathbf{A}_k - (\mathbf{D}_k^\top\mathbf{X}_k - \mathbf{Z}_{1,k})\|_F^2 + \lambda_2\|\mathbf{A}_k\|_{w,*}, \tag{4}$$

$\mathbf{A}_k$ can be solved with weighted nuclear norm (WNN) [9]

Updating $\{\{\mathbf{A}_{\overline{k}} : \overline{k} \in r(k)\}\}$: Given fixed $\{\mathbf{D}, \mathbf{A}, \mathbf{Z}\}$, we solve the l_1 norm minimization problem over $\overline{\mathbf{A}}_k$:

$$\underset{\{\mathbf{A}_{\overline{k}}:\overline{k}\in r(k)\}}{\arg\min} \sum_{\overline{k}\in r(k)} \left(\|\mathbf{A}_{\overline{k}} - (\mathbf{D}_k^\top\mathbf{X}_{\overline{k}} - \mathbf{Z}_{2,\overline{k}})\|_F^2 + \lambda_1\|\mathbf{A}_{\overline{k}}\|_1\right) \tag{5}$$

which can be solved with element-wise soft-thresholding:

Algorithm 1: The proposed Algorithm

Input: The training features of calcium image and behavior image $\mathbf{X}_k$, k = 1, . . . , K and spike group as the label for supervised learning;

Output: The analysis dictionary $\mathbf{P}_k \in \mathbb{R}^{M\times S}$, synthesis dictionary $\mathbf{D}_k \in \mathbb{R}^{S\times M}$ and the regression coefficients $\boldsymbol{\beta}$.

Set $\mathbf{A}_k = \mathbf{P}_k\mathbf{X}_k$,$\mathbf{A}_{\overline{k}} = \mathbf{P}_k\mathbf{X}_{\overline{k}}$ for $\overline{k} \in r(k)$ and $\mathbf{Z}_{1,k} = \mathbf{Z}_{2,k} := 0$, $k = 1, \ldots, K$;

while *not converged* **do**

  Find groups of each feature with same label k;
  Update $\mathbf{D}_k$ using Eq. (3);
  Update $\mathbf{A}_k$ by solving Eq. (4);
  Update $\{\{\mathbf{A}_{\overline{k}} : \overline{k} \in r(k)\}$ by solving Eq. (5);
  Update Lagrange multiplier $\mathbf{Z}_k$ using $\mathbf{Z}_{1,k} := \mathbf{Z}_{1,k} + (\mathbf{A}_k - \mathbf{P}_k\mathbf{X}_k)$ and $\mathbf{Z}_{2,k} := \mathbf{Z}_{2,k} + (\overline{\mathbf{A}}_k - \mathbf{P}_k\overline{\mathbf{X}}_k)$.;

return x ;

$$\{\{\mathbf{A}_{\overline{k}} : \overline{k} \in r(k)\}\}_i = \sum_{\overline{k}\in r(k)} \left(\mathbf{sign}\big([\mathbf{D}_k^\top \mathbf{X}_{\overline{k}} - \mathbf{Z}_{2,k}]_i\big) \cdot \left([\mathbf{D}_k^\top \mathbf{X}_{\overline{k}} - \mathbf{Z}_{2,k}]_i - \tfrac{\lambda_1}{\mu_2}\right)_+ \right) \tag{6}$$

where, $(\cdot)_+$ is defined as $(x)_+ = x, x \geq 0$ and $(x)_+ = 0$ for others to any x.

In the K-th step, a linear regression model was furter learned for the minority class as

$$\mathbf{y} = \mathbf{X}_K\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{7}$$

where, y is the label in the class K, $\boldsymbol{\beta}$ is the learned regression coefficients, $\boldsymbol{\varepsilon}$ is noise term.

Finally, we update the dual variables with $\mathbf{Z}_{1,k} := \mathbf{Z}_{1,k} + (\mathbf{A}_k - \mathbf{P}_k\mathbf{X}_k)$ and $\mathbf{Z}_{2,k} := \mathbf{Z}_{2,k} + (\overline{\mathbf{A}}_k - \mathbf{P}_k\overline{\mathbf{X}}_k)$.

In this paper, we set $\mu_1 = 200$, $\lambda = 0.1$, $\lambda_1 = 0.001$ and $\lambda_2 = 0.0001$ and dictionary size $\mathbf{D}$=30.

II-C. Classification

Let $\mathbf{x}^i \in \mathbb{R}^{S_i}$ be the features of modality i to classify. We define $e_k^i = \|\mathbf{x}^i - \mathbf{D}_k^i {\mathbf{D}_k^i}^\top \mathbf{x}\|_2$ as the error of $\mathbf{x}^i$ with the dictionaries of class k for feature type i. The class k of the sample is decided based on the $\mathbf{D}_k$ that gives the smallest error as follows,

$$\widehat{k}_i = \arg\min_k e_k^i \tag{8}$$

Fusion impactor α_i: To combine the information of feature from multi-modalities with optimized weight $\boldsymbol{\alpha}$, we use the validation subset to learn a regression model as follows.

$$\arg\min_{\boldsymbol{\alpha}} \ \left(k_{\text{real}} - \sum_i \alpha_i\,\widehat{k}_i\right)^2, \ \text{ s.t. } \sum_i \alpha_i = 1, \ \alpha_i \geq 0, \forall i. \tag{9}$$

Constraints on regression coefficients α_i enforce the final prediction to be a convex combination of predicted values for each feature type. i.e., $k_{\text{predicted}} = \sum_i \alpha_i\,\widehat{k}_i$, $i = 2$

Table I: The total amount of spike in each spike group.

Group	**1**	**2**	**3**	**4**
#	25390	3168	1523	339
Group	**5**	**6**	**7**	**8**
#	69	15	1	2

In the case of predicted class is K, we also apply the trained regression coefficients $\boldsymbol{\beta}$ to furter predict the minarity classes.

III. EXPERIMENTS

III-A. Dataset

The proposed framework is evaluated on modeling single neuron spiking activity with calcium imaging and behavior video of 30507 slides for each one. For the dataset, a 30-second sliding window (step size: 15 seconds) is applied to the spike train to generate a number of sequences. For each sequence, the firing rate and power spectrogram are computed. The average power is computed the average magnitude of the power spectrogram. The distribution of spikes is presented in Fig. 2 a) and the number of spikes in each class is shown in Table I. For the detailed about the dataset collection, please refer to the materials in [5].

III-B. Metrics

The 10-fold cross-validation is applied on this experiments. To measure performance in terms of prediction overall accuracy (oACC), balanced accuracy(bACC), a root mean square error (RMSE) and mean absolute error (MAE) are applied in this paper. $oACC = \dfrac{N_c}{N_t}$, where, N_c is total number of all correctly classified subjects and N_t is number of all test subjects. $bACC = \frac{1}{K}\sum_{k=1}^{K} \dfrac{N_c^k}{N_t^k}$, where, N_c^k is total number of all test subjects in class k. N_t^k is total number of all test subjects in class k. The experiments were conducted with Matlab R2017b using a i7-6700K CPU with 16GB of RAM.

III-C. Prediction of neuron spiking

The performance of the proposed approach is demonstrated by predicting the group-patch based single-neuron spiking, based on calcium imaging and behavior video. Here, we evaluate our method in a classification setting by applying a sliding window on the 30507 spikes as of figure 3 the data preprocessing scheme.

Table II compares the RMSE, MAE, oACC and bACC obtained by our approach to the recently proposed method NDPL[10], variational autoencoders(VAE) and random forest (RF)[12] for evaluating the proposed model and the impact of each modality in the model. We see that the proposed approach outperforms the NDPL method. OnB is our model

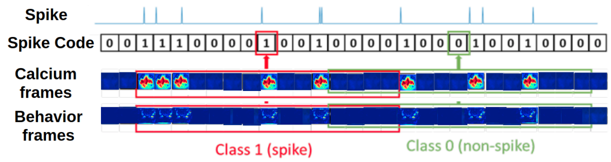

Fig. 3: Data preprocessing scheme

Table II: Classification results on the single neuron spikes with 5-fold CV.

	RMSE	MAE	oACC	bACC
RF	0.8332	0.5034	0.5680	0.2300
NDPL	0.6688	0.2685	0.8692	0.3124
VAE	0.5182	0.4700	0.7371	0.2718
OnB(DL)	0.7451	0.4317	0.6340	0.2881
OnI(DL)	0.7339	0.3864	0.6977	0.1228
Ours(DL)	0.5017	0.2353	0.7930	0.3963

OnB: only behavior frame, OnI: only calcium imaging, DL:Dictionary Learning, VAE: Variational AutoEncoder

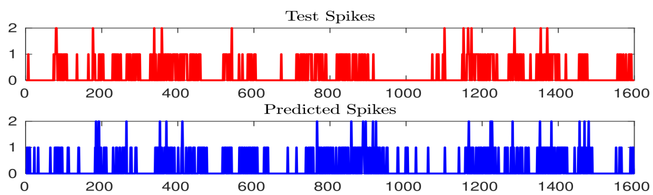

Fig. 4: Data preprocessing scheme

only with a behavior video frame as input, and similarly OnI is the proposed model only with calcium image as input features. From Table II, we can find the behavior frame as input features have higher residual classification errors (RMSE and MAE), compared with only the calcium image features as input. With both calcium image features and behavior video frame features as input, the proposed achieved the best performance, yielding improvements of about 0.1671 in RMSE, 0.0332 in MAE and 0.0839 in bACC. Figure 4 is one example of predicted spikes with the proposed approach and the test spikes.

The classification fusion performance with the variables (i.e., calcium image, behavior video frame) is also evaluated in this section. The fusion weight for behavior video frame and calcium image is $\alpha_b = 0.5415$ and $\alpha_I = 0.4584$, respectively. It is consistent with the fact that the correlation of spikes and principal components of calcium image is $r_{si} = 0.0979$, and the correlation of spikes and principal components of behavior is $r_{sb} = 0.2478$, as shown in figure 2.

IV. CONCLUSION

A group-patch based classification and the asymptotic prediction model is proposed for predicting and modeling neuron spikes with calcium image and behavior video frame. This approach learns subspace discriminative features by imposing tightest robust principal component analysis with low-rank constraints on projective coefficients, and a l_1 sparsity constraint on coefficients of non-class. To take the local similar features into consideration, this paper proposed an overlapped patch approach, and general classification guided asymptotic prediction for the minority spikes in the imbalanced dataset. Experiments showed the benefit of our approach. In the future, we would use truncated-correlation photothermal coherence tomography derivative imaging modality technique [13] to achieve more multi-modality data.

ACKNOLEDGEMENTS

This work was supported, in part, by the Fonds de recherche du Quebec (CCC 246110, 271636), National Nature Science Foundation of China (NSFC: 61902220, 61602277, U1609218).

J.-B.P. was partially funded by National Institutes of Health (NIH) NIH-NIBIB P41 EB019936 (ReproNim) NIH-NIMH R01 MH083320 (CANDIShare) and NIH RF1 MH120021 (NIDM), the National Institute Of Mental Health of the NIH under Award Number R01MH096906 (Neurosynth), as well as the Canada First Research Excellence Fund, awarded to McGill University for the Healthy Brains for Healthy Lives initiative and the Brain Canada Foundation with support from Health Canada. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or other funding agencies.

National Institute Of Mental Health of the National Institutes of Health under Award Number R01MH096906. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

V. REFERENCES

[1] Ari S Benjamin, Hugo L Fernandes, Tucker Tomlinson, Pavan Ramkumar, Chris VerSteeg, Lee Miller, and Konrad P Kording, "Modern machine learning far outperforms glms at predicting spikes. biorxiv," 2017.

[2] Carsen Stringer, Marius Pachitariu, Nicholas Steinmetz, Charu Bai Reddy, Matteo Carandini, and Kenneth D Harris, "Spontaneous behaviors drive multidimensional, brainwide activity," *Science*, vol. 364, no. 6437, pp. 255–255, 2019.

[3] Wilson Truccolo, Leigh R Hochberg, and John P Donoghue, "Collective dynamics in human and monkey sensorimotor cortex: predicting single neuron spikes," *Nature neuroscience*, vol. 13, no. 1, pp. 105, 2010.

[4] Hiroshi Makino, Chi Ren, Haixin Liu, An Na Kim, Neehar Kondapaneni, Xin Liu, Duygu Kuzum, and Takaki Komiyama, "Transformation of cortex-wide emergent properties during motor learning," *Neuron*, vol. 94, no. 4, pp. 880–890, 2017.

[5] Dongsheng Xiao, Matthieu P Vanni, Catalin C Mitelut, Allen W Chan, Jeffrey M LeDue, Yicheng Xie, Andrew CN Chen, Nicholas V Swindale, and Timothy H Murphy, "Mapping cortical mesoscopic networks of

single spiking cortical or sub-cortical neurons," *Elife*, vol. 6, pp. e19976, 2017.

[6] Mingli Zhang, Yuhong Guo, Caiming Zhang, Jean-Baptiste Poline, and Alan Evans, "Modeling and analysis brain development via discriminative dictionary learning," in *International Workshop on Machine Learning for Medical Image Reconstruction*. Springer, 2019, pp. 80–88.

[7] Simon Kornblith and Doris Y Tsao, "How thoughts arise from sights: inferotemporal and prefrontal contributions to vision," *Current Opinion in Neurobiology*, vol. 46, pp. 208–218, 2017.

[8] Mingli Zhang, Christian Desrosiers, Yuhong Guo, Caiming Zhang, Budhachandra Khundrakpam, and Alan Evans, "Brain status prediction with non-negative projective dictionary learning," in *International Workshop on Machine Learning in Medical Imaging*. Springer, 2018, pp. 152–160.

[9] Mingli Zhang and Christian Desrosiers, "High-quality image restoration using low-rank patch regularization and global structure sparsity," *IEEE Transactions on Image Processing*, vol. 28, no. 2, pp. 868–879, 2018.

[10] Mingli Zhang, Christian Desrosiers, Yuhong Guo, Budhachandra Khundrakpam, Noor Al-Sharif, Greg Kiar, Pedro Valdes-Sosa, Jean-Baptiste Poline, and Alan Evans, "Brain status modeling with non-negative projective dictionary learning," *NeuroImage*, pp. 116–226, 2019.

[11] Shai Shalev-Shwartz and Shai Ben-David, *Understanding machine learning: From theory to algorithms*, Cambridge university press, 2014.

[12] Tong Tong, Qinquan Gao, Ricardo Guerrero, Christian Ledig, Liang Chen, Daniel Rueckert, Alzheimer's Disease Neuroimaging Initiative, et al., "A novel grading biomarker for the prediction of conversion from mild cognitive impairment to Alzheimer's disease," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 1, pp. 155–165, 2017.

[13] Hai Zhang, Pantea Tavakolian, Koneswaran Sivagurunathan, Andreas Mandelis, Wei Shi, and Fei-Fei Liu, "Truncated-correlation photothermal coherence tomography derivative imaging modality for small animal in vivo early tumor detection," *Optics letters*, vol. 44, no. 3, pp. 675–678, 2019.