Evento	Salão UFRGS 2019: SIC - XXXI SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
Ano	2019
Local	Campus do Vale - UFRGS
Título	Segmentation of Ultrasonic Vocalizations in Infant Mice using Deep Learning
Autor	GUSTAVO MADEIRA SANTANA
Orientador	SERGIO BAMPI

Segmentation of Ultrasonic Vocalizations in Infant Mice using Deep Learning

Autor: Gustavo Madeira Santana. Orientador: Sergio Bampi
Instituto de Informática - Universidade Federal do Rio Grande do Sul

Infant vocalizations are important for development and survival. Specifically, they seem to share the same function of attracting caregiver attention across mammalian species. Problems with infant vocalization, such as speech disorders, are typically the first symptoms of serious neurodevelopmental disorders; making it an important point for early diagnosis and intervention. Here, we use mice as an animal model to study vocalizations. Impaired vocalization in mice can lead to maternal neglect and infant´s death. Therefore, studying vocalizations can give unique insights into infant mice behavior and mother-pup interactions. Vocalization in mice is believed to be predominately inherited. Our goal is to identify vocal features that change when rearing litters of different strains in standard and foster conditions, showing that external factors might influence the vocal behavior. Having a robust method to analyze vocalizations becomes essential. We devised a fully-convolutional semantic segmentation network to analyze spectral features of mouse ultrasonic vocalizations recordings under experimental conditions.

To assess maternal contributions to vocal behavior, we cross-foster two distinct mouse strains – C57BL/6J (B6) and A/J. Mice used in the experiments were ten days old from both genders. For cross-fostering rearing, half of the litters were exchanged between mothers within the first 24 hours after birth of the litter. Ultrasonic vocalizations were recorded using the UltraSoundGate Condenser Microphone CM16. Each animal is placed inside an isolation chamber with fresh bedding and microphones are placed 10 cm above the mice. The segmentation network and supporting scripts were implemented in Python v3.6 using the deep learning framework PyTorch.

Ultrasonic vocalizations recordings are processed using signal- and image-processing techniques to identify candidate vocalizations in the recordings. The pipeline consists of a short-time Fourier Transform to obtain spectrograms; series of morphological operations on the spectrograms; and post-processing to differentiate vocalizations from noise. The vocal identifier outputs spectrograms of the candidate vocalizations. These candidates are used as input for the segmentation network which outputs a precise mask of the vocalization in the frequency domain. Spectral features of the vocalizations, such as mean frequency, bandwidth, duration, and intensity are all saved to a table for posterior analysis.

The segmentation network is based on the encoder-decoder paradigm. The contracting encoder path keeps the spatial segmentation information while sacrificing details such as edges. The decoder path up-samples the segmentation back to the original size of the input image, and by using skip-connections from the decoder path is able to recover the lost fine-grain details to reconstruct the final segmentation. Every block consists of a series of Convolution-BatchNormalization-ReLU and max-pooling layers. Encoder blocks use convolution layers for down-sampling. Decoder blocks use transposed convolution layers for up-sampling. The loss function for training is a weighted sum of the binary cross-entropy loss and the Dice coefficient, and training is performed using stochastic gradient descent.

The network achieves mean Intersection over Union score of 98.82% when compared to the ground truth. Preliminary results show that mouse pups reared with a mother of a different strain can alter their vocalization behavior (e.g., mean frequency), showing that some vocalization features might be plastic and not entirely inherited.