**Deep Learning-Based High-Resolution Time Inference for Deciphering Dynamic Gene Regulation from Fixed Embryos**

**1 Introduction**

The code package consists of two sections: one designed for analyzing confocal image stacks of fixed *Drosophila* embryos and the other for establishing the time inference framework. This document provides the essential information needed to use the code. For additional technical details, please refer to the **Methods and Supplementary Note**.

The code requires pre-installation of MATLAB with the image analysis, curve fitting, optimization, and parallel processing toolboxes, as well as the Bio-Formats package (available at [Bio-Formats download](https://www.openmicroscopy.org/bio-formats/downloads/)). Cellpose is also needed for nuclei segmentation (installation guide at [Cellpose GitHub](https://github.com/MouseLand/cellpose)). For model training and prediction within our time inference framework, the Python-based deep learning API Keras must be installed ([Keras website](https://keras.io/" \t "_new)). Here we use CUDA 11.2 and cuDNN 8.1.1 for GPU acceleration, with Python 3.7 managed through Anaconda. TensorFlow 2.7.0, which includes Keras, is installed.

For optimal performance, we recommend using a 64-bit computer system with a multi-core processor (≥36 cores), large RAM (≥256 GB), and a high-performance GPU (≥16 GB). To install the package, simply copy the folder “FISHIF-Time” to your computer and add “FISHIF-Time/MATLAB” to the MATLAB search path.

**2 Image analysis of fixed images labeled by smFISH and immunofluorescence**

**2.1 Image conversion and preprocessing**

Before running the code, go to the folder “FISHIF-Time/Confocal”, and create a subfolder (e.g. “Exp1”) to store your confocal image files. A sample image file is available at http://gofile.me/4yuzx/1vxvuoUCF. In addition, download folder “Standard” into the “FISHIF-Time” folder (available at <http://gofile.me/4yuzx/3MbMTb7H2>). Next, in the folder “FISHIF-Time”, find the Excel file “Duallist.xls”. Open the file, and add the information regarding your data using the following format:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Folder name** | **Leave Blank** | **Channel index** | **Tiling method** | **Channel name** | **Running index** |
| Confocal/Exp1/ |  | [4,4,3,2,1] | [2,1;2,2] | {'A647','TMR','A488','A405'} | T |

After saving the Excel file, run the following scripts:

1. To convert the confocal LSM files into TIF images, run “cziconvert2('Duallist.xls');”.
2. To correct for optical aberrations, run “flyimagetile4\_quick('Duallist.xls',[],[],1);”.
3. To create a results folder and save basic information for further analysis, run “Dualprocess0;”.
4. For manual refinement of the embryo segmentation, run the MATLAB GUI “emmask\_mannual;”.

Once done, you will have a subfolder called “stacks” inside your data folder (e.g. “Exp1”) containing all TIFF files, and another subfolder “Results” for future analysis results.

**2.2 Nuclear segmentation**

To reconstruct the 3D shape of individual nuclei in the embryo, run the following scripts:

1. For automatic segmentation, run the MATLAB APP “NuclearRecognitionProcess”. After loading the path “FISHIF-Time/Duallist.xls”, first run “try” for quick segmentation. You need to input the signal channel for nuclear segmentation and resize ratio. Once done, a subfolder called “raw2input\_try” will be created inside “stacks” subfolder, containing the quick segmentation result file “time0001Try\_seg.npy”. You can open and check in the Cellpose GUI. Adjust the resize ratio and retry if needed to obtain a satisfactory quick segmentation result. After achieving a satisfactory “try” result, click “Automatic Operation” to generate the 3D segmentation result. You can check the “Edge” option to ensure that all recognized nuclei are within the embryo region. Depending on the speed of your computer, this step may take a few hours.
2. For manual refinement of the automatic segmentation result, run the MATLAB GUI “nuclei\_manual3D\_foci”.

All segmentation results are stored in the subfolder “masks” inside your data folder.

**2.3 Time inference**

For automatic time inference, go to the folder “FISHIF-Time/Time inference APP” and run the MATLAB APP “EmbryoTimePredict”. Additionally, download files in the “FISHIF-Time/Trained CNN models/DNA-based models” folder and place it inside the “FISHIF-Time/Time inference APP” folder before running the code. After loading the data folder path “Confocal/Exp1/” and specifying the installed path of Keras, click “Automatic Operation” to infer the developmental time. The default value of “Standard area” is tailored for OreR *Drosophila* embryo. If applying the program to other strains, you must input the corresponding value. This step may take tens of minutes per embryo, depending on your computer's speed. For a faster but approximate result, you can select the “Fast” mode.

All time inference results are stored in the Excel file “out.xls” inside your data folder. The final inference result and corresponding standard deviation (SD) for each embryo are shown in the last two columns of the table.

**2.4 smFISH spot analysis**

To identify smFISH spots and quantify their intensities, run the following scripts:

1. For automatic spot quantification, run the MATLAB GUI “stack\_RNA”. Nascent and cytoplasmic mRNAs need to be identified separately by choosing different “Analysis types” (“foci” or “single”) in the GUI.
2. To extract the typical intensity of a single mRNA, run “hist\_fit\_RNA('Duallist.xls');”.
3. To double-check and improve the identification of nascent mRNAs, run the MATLAB GUI “stack\_RNA\_check”.

Results of nascent and cytoplasmic mRNAs are stored in the subfolders “Histogram” and “Histogram\_A” respectively, inside your data folder.

**2.5 Immunofluorescence (IF) spot analysis**

To identify IF spots and quantify their intensities, run the following scripts:

1. For automatic spot quantification, run the MATLAB GUI “stack\_protein\_new4”. Anterior, medial, and posterior spots need to be identified separately by choosing different “Analysis types” (“Anterior”, “Medial”, or “Posterior”) in the GUI.
2. To extract the typical intensity of a single protein molecule, run “hist\_fit\_protein('Duallist.xls');”.

Results for anterior, medial, and posterior IF spots are stored in the subfolders “Histogram\_protein\_A”, “Histogram\_protein\_M” and “Histogram\_protein\_P” respectively, inside your data folder.

**2.6 Analysis of transcriptional regulation**

To analyze transcriptional regulation, run “Dualprocess7”. The key results are saved as a MAT file, with the same name as your image file, and stored in the subfolder “Results” inside your data folder.

**2.7 Reconstruction of time sequence data**

After completing **steps 2.1–2.6**, the quantitative analysis and time inference for each fixed embryo will be done. Time-resolved datasets for each cycle can then be reconstructed by integrating data from multiple embryos. In the folder “FISHIF-Time/Confocal”, find the Excel file “Fit.xlsx” with three sheets “cycle11”, “cycle12”, and “cycle13”. Open the file, and add the information regarding your data using the following format:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Path** | **Folder name** | **Embryo name** | **Embryo index** | **Protein label** |  |  |  | **Time** | **RNA channel** |
| FISHIF-Time/Confocal/ | Exp1 | 20210311-kr-r-TMR-Bcd-p-647-hb-p-488-Hoch405-oreR-s2\_152\_5 | 1 | H |  |  |  | 5.36106 | 2 |

**2.8 Fitting the spatiotemporal relationship between RNA and its TFs**

To fit the nuclear mRNA levels in response to TF concentrations from time-resolved datasets, run “plot\_in\_nc11\_fit”, ” plot\_in\_nc12\_fit”, “plot\_in\_nc13\_fit” for nc11-13, respectively. Comments in these m-files provide detailed description of the inputs and outputs of the functions.

**2.9 Fitting the nascent mRNA distribution**

To fit the unsteady-state nascent mRNA distribution to a time-dependent two-state model with *P*active, go to the folder “FISHIF-Time/simu\_FISH” and run “TimeFitMain”. Comments in this m-file provide detailed description of the inputs and outputs of the functions.

**3 Establishment of the time inference framework**

**3.1 Live image conversion and preprocessing**

Before running the code, go to the folder “FISHIF-Time/Confocal”, and create a subfolder (e.g. “LiveExp1”) to store your confocal image files. A sample image file is available at http://gofile.me/4yuzx/1vxvuoUCF. Next, in the folder “FISHIF-Time/Confocal”, find the Excel file “Livedata.xlsx” with three sheet “cycle11”, “cycle12”, and “cycle13”. Open the file, and add the information regarding your data using the following format:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Folder name** | **File name** | **Duration** | **Mode** | **Running index** | **Embryo index** | **Z layer** | **Start time** | **Size** |
| LiveExp1 | Image 1 | 10 | Confocal | T | 1 | 21 | 4 | 512 |

After saving the Excel file, run the following scripts:

1. To convert the confocal LSM files into TIF images, run “lsmtiff”.
2. To generating cropped images, run “DAPI\_time”. You can adjust “DAPI\_th” and retry to get a satisfied nuclear segmentation mask image to ensure that the output cropped images are appropriate.

Comments in these m-files provide detailed description of the inputs and outputs of the functions.

**3.2 Training and testing Histone-based CNN models**

Before running the code, install “jupyter notebook” in your keras environment, and then run the following scripts:

1. For training Histone-based CNN models, run “regression\_cycle\_kfold-Histone”.
2. For testing CNN models in live images, run “regression\_cycle\_time.predict-Histone”.

Comments in these ipynb-files provide detailed description of the inputs and outputs of the functions.

**3.3 Fixed image conversion and preprocessing**

Before running the code, go to the folder “FISHIF-Time/Confocal”, and create a subfolder (e.g. “FixedExp1”) to store your confocal image files. A sample image file is available at http://gofile.me/4yuzx/1vxvuoUCF. Next, in the folder “FISHIF-Time”, find the Excel file “Duallist.xls”. Open the file, and add the information regarding your data using the following format:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Folder name** | **Leave Blank** | **Channel index** | **Tiling method** | **Channel name** | **Running index** |
| Confocal/FixedExp1/ |  | [3,3,2,1,1] | [2,1;2,2] | {'TMR','TMR',’A405’} | T |

After saving the Excel file, run the scripts in **Section 2.1** and perform the automatic nuclear segmentation described in **Section 2.2**.

Next, integrate multiple embryos based on nuclear cycles. In the folder “FISHIF-Time/Confocal”, find the Excel file “Fixeddata.xlsx” with three sheets “cycle11”, “cycle12”, and “cycle13”. Open the file, and add the information regarding your data using the following format:

|  |  |  |  |
| --- | --- | --- | --- |
| **Folder name** | **Embryo name** | **Embryo index** | **Nuclear**  **diameter** |
| FixedExp1 | 20210403-23651-hisRFP-TMR-Hoch405-5\_159\_002 | 1 | 45.72481 |

After saving the Excel file, run “DAPI\_fixed\_his” to generate cropped images for the anterior, medial, and posterior regions. Adjust “DAPI\_th” as needed and rerun the script to obtain a satisfactory nuclear segmentation mask, ensuring that the output cropped images are appropriate. Use the parameter “scale” to modify the cropped image size. All cropped images are saved in a subfolder named after the corresponding “scale” value within the “DAPI” folder.

**3.3 Time inference from Histone images with size rescaling**

To predict time from Histone images of fixed embryo, run the following scripts:

1. For time inference by CNN models in fixed image, run “regression\_cycle.predictlist”.
2. Before running the code, open the Excel file “Fixeddata.xlsx” and add a column regarding the inferred time (Fifth row) corresponding to your data. To estimate the scaling magnitude, run “plot\_scale\_D”.

Comments in these files provide detailed description of the inputs and outputs of the functions.

Using the updated magnitude, rerun “DAPI\_fixed\_his” from **section 3.2** and the scripts in **section 3.3**. Repeat this process approximately three times until the estimated magnitude matches the input magnitude in “DAPI\_fixed\_his”. Note that the inferred times must be updated in each iteration.

**3.4 Relayed learning for DNA-based CNN models**

1. To generate cropped DNA images for training and testing DNA-based CNN models, run “DAPI\_fixed\_hoch”. Adjust “DAPI\_th” as needed and rerun the script to obtain a satisfactory nuclear segmentation mask, ensuring that the output cropped images are appropriate. Cropped images for the anterior, medial, posterior, and whole embryo regions are saved in in separate subfolders within “DAPI”, such as “DAPI\_Hoch\_s\_Aw”, “DAPI\_Hoch\_s\_Mw”, “DAPI\_Hoch\_s\_Pw”, and “DAPI\_Hoch\_s\_w” for small spatial scale datasets.
2. To calibrate image contrast and saturation for DNA images, run “datagenerate2\_noise”.
3. For training DNA-based CNN models, run “regression\_cycle\_kfold-DNA”.
4. For testing DNA-based CNN models in fixed image, run “regression\_cycle.predictlist”.

Comments in these files provide detailed description of the inputs and outputs of the functions.