

1. Provide your team background and organization description (if applicable).

My name is Xueer Chen. I am the sole member of my team. I am a compute bio researcher from Bristol Myers Squibb. My research interests are single-cell multi-omics and drug discovery.

2. Explain why you participated in the Cytologia challenge.

The Cytologia Challenge presented an opportunity to learn cell image analysis, a field that directly intersects with my research interests. As a single-cell researcher, I am constantly seeking ways to enhance my understanding of cellular behavior and function. The challenge provided a platform to not only test my existing knowledge but also to potentially develop novel approaches to cell image understanding.

Furthermore, participating in this competition allowed me to engage with a community of like-minded researchers and professionals who share a passion for this field. The exchange of ideas and perspectives is invaluable for advancing scientific discovery.

3. Describe how you built your winning model and elaborate on the technical and modeling choices you made.

My approach is the ensemble of 16 models, consisting of 4 detection and 12 classification models. The key is to build diversified models with proper augmentations and ensemble by averaging their probabilities.

please refer to the technical report for details.

https://docs.google.com/document/d/1uaP4lo6_42sqa-3KeCuWS1XZ6RbSs66OVbKWH9cj0iM/edit?usp=sharing

model weights: https://drive.google.com/drive/folders/1t-airGVEiwAGD20PRHKgLZ7NKLPxyDIX?usp=drive_link

colab inference notebook:

<https://colab.research.google.com/drive/1tPdwVhFCnJJAZdAWC49BVoAgfv0FIPD0?usp=sharing>

github repo: https://github.com/xueerchen1990/cytologia_2nd_place

4. What GPU/CPU/RAM resources you used to build your model

local workstation:

- one RTX 3090 GPU 24GB GPU memory

- AMD Ryzen 9 9950X CPU
- CPU memory 32GB

5. Do you have any positive feedback or improvement opportunities for the Trustii.io platform?

- The hosts of the forum are highly responsive and ensure that all questions are answered in a timely manner.
- Additionally, the data provided is well-prepared and organized, making it easy to access and utilize.