Orals

Jiangmei (Ruby) Xiong

2023-08-15

Table of contents

Statistical Analysis in Multiplexed Immunofluorescence Imaging

This is the document for doctoral oral exam of Jiangmei Xiong. The topic of this oral exam is data preprocessing of mIF imaging data. This document will walk you through the basics of multiplexed immunofluorescence image, Jiangmei's first dissertation paper, and a literature review for missing data imputation in multiplexed immunofluorescence imaging.

1 Chapter 1: Introduction to Multiplexed Immunofluorescence Images

1.1 Multiplexed Immunofluorescence Images

Multiplexed Immunofluorescence (mIF) Image is a recent development from Immunofluorescence (IF), a branch of Immunohistochemistry (IHC). The first structural conceptualization of IHC is established in 1941. Coons, Creech, and Jones (1941) described that in formalin-prefixed mammalian tissue, there is a type of antibody that can be identified by fluorescent antigens. Since then, IHC is developed into an important tool for cancer diagnosis (Duraiyan et al. 2012). Within the word "immunohistochemistry", "immuno" refers to the antigen-antibody reaction in the process, "histo" means tissue, and "chemistry" is the process. During IHC, antibody can be tagged with labels such as enzyme, fluorochromes, which reacts when the corresponding antigen-antibody bind is formed (Ramos-Vara 2005). Similarly, the word immunofluorescence can split into "immuno" and "fluorescence", and "fluorescence" corresponds to the fluorescent signal generated by the fluorocromes(Hussaini, Seo, and Rich 2022).

IHC/IF can only detect one biomarker for a tissue region. This limitation makes IHC/IF unable to identify more complicated expression patterns that require more than one biomarker (Sheng et al. 2023). The development of multiplexed IHC (mIHC)/IF(mIF) image resolved this issue. Multiplexed IHC/IF image display different protein information for each plex of the image, while retaining the spatial and morphological information of the tissue (Eng et al. 2022). mIHC/mIF can be seen as a stack of images, each presenting a different portion of the same tissue.

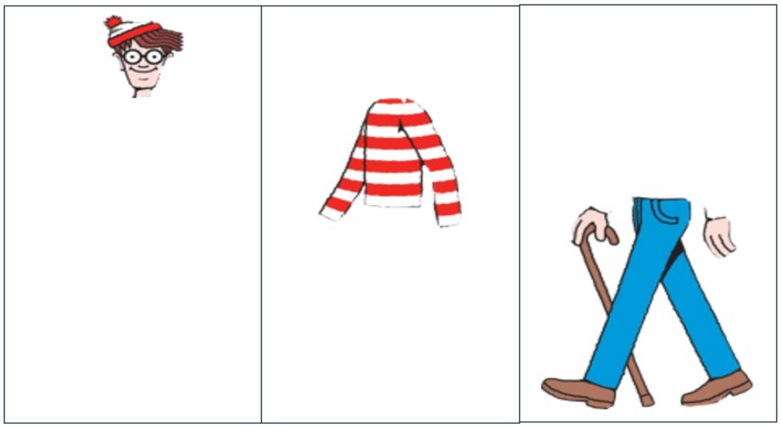


Figure 1.1: This is what it might look like if Waldo is to be analyzed in mIF style (?Maybe?).

Figure 1.2 by Sheng et al. (2023) shows several different methods for creating mIF images. It can be seen that the first two methods both uses cycles of stain-photo-removal, and the last method is a one-off step where all labels are tagged at once. Sheng et al. (2023) also tabulated all multiplexd IHC/IF technologies, where the number of biomarkers that can be identified ranges from 4 to 100.

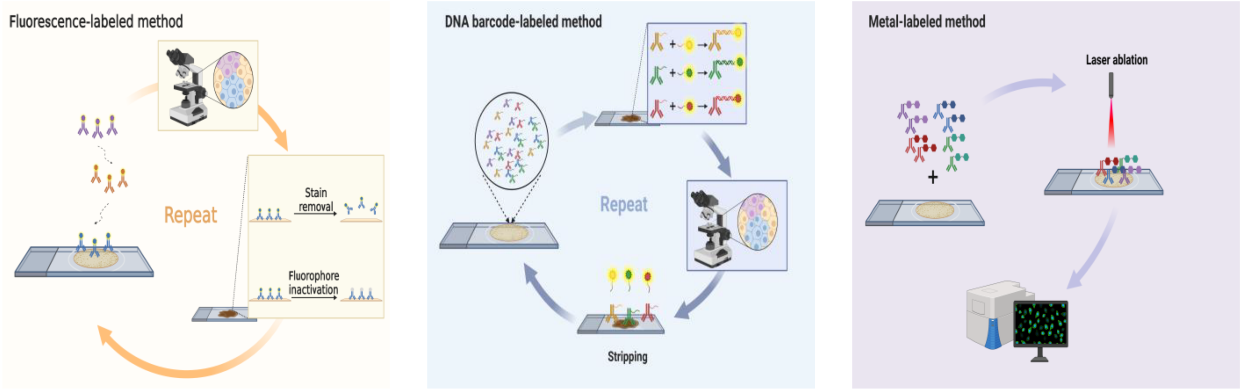


Figure 1.2: Different methods for creating mIF/mIHC images. Image courtesy of Sheng et al. (2023)

mIHC/mIF images are widely used in studies for immune tumor microenvironments (iTME). The studies often involves cell type proportions within a certain region, spatial clustering of immune cells or spatial interaction among different cell types (Julia Wrobel, Harris, and Vandekar 2023). For example: Schürch et al. (2020) discovered that within granulcyte cell meightborhood, the enrichment PD-1+CD4+ T cells are correlated with the survival outcome of a subset of colorectal cancer patients; B. Chen et al. (2021) shows different immune cell proportion and clustering between different colorectal tumor types; Steinhart et al. (2021) found that certain immune cell proportions and spatial interactions are correlated with ovarian cancer patient survival outcomes.

1.2 Data Structure

For mIHC/mIF images, each individual "plex" corresponds to a different immune protein identified by a type of stain. Each plex goes through analogous transition as shown in Figure 1.3 and form a table of cell expressions, and the tables of different protein expressions are combined in the end.

Initially, greyscale intensity is assigned to each pixel. The greyscale intensity is taken as the intensity of marker expression. The image then goes through cell segmentation, which are usually based on machine learning or deep learning methods(McKinley et al. 2022; Schüffler et al. 2015). DAPI, a fluorescent stain typically used for cell morphology identification, is most often used for the initial cell segmentation. The same cell-segmentation will be used for all other marker channels. Next, the pixel intensities, pixel positions, and the cell that the pixel belongs to are entered into a dataframe. Finally, the pixel intensities and position will be averaged within the cell group. Often, the median of pixel intensities is used as well, to reduce the impact of pixel intensity outliers. The end result after combining all marker channels will be a dataset with each row representing an individual cell, columns of different marker expressions, and cell properties such as position, cell type (e.g. tumor cell or not, tissue type).

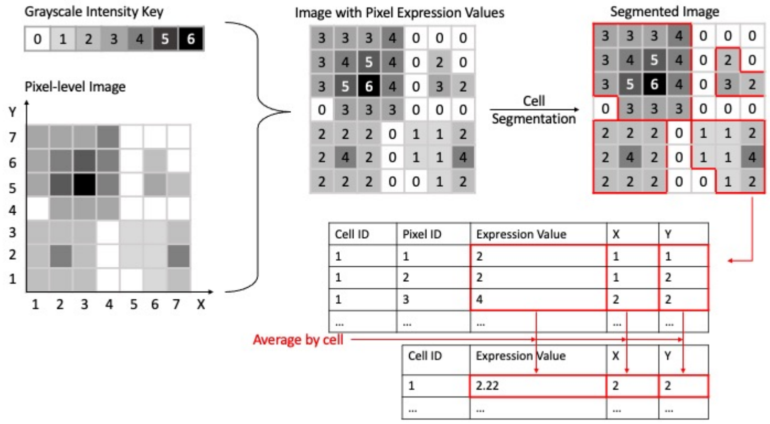


Image with Pixel Expression Values

3	3	3	4	0	0	0
3	4	5	4	0	2	0
3	5	6	4	0	3	2
0	3	3	3	0	0	0
2	2	2	0	1	1	2
2	4	2	0	1	1	4
2	2	2	0	0	1	2

Segmented Image

3	3	3	4	0	0	0
3	4	5	4	0	2	0
3	5	6	4	0	3	2
0	3	3	3	0	0	0
2	2	2	0	1	1	2
2	4	2	0	1	1	4
2	2	2	0	0	1	2

Cell ID	Pixel ID	Expression Value	X	Y
1	1	2	1	1
1	2	2	1	2
1	3	4	2	2
...	...	...	...	...

Cell ID	Expression Value	X	Y
1	2.22	2	2
...	...	...	...

Figure 1.3: Transition from a single-plex mIHC/mIF image to a single-cell dataframe. The greyscale intensity range is only a demonstration. In application, the range of the greyscale intensity depends on the configuration of image softwares. The cell intensity of the bottom left cell is shown in the table as an example.

2 Chapter 2: GammaGateR

2.1 Research question

Before important spatial insights can be gleaned using statistical methods, mIF images undergo an intensive preprocessing pipeline to obtain single-cell measurements. While there are various steps included in the pipeline such as image registration, single-cell segmentation, quantification, and batch correction, cell phenotyping is typically the final step before downstream analyses on the cell-level data (Graf et al. 2022; Harris et al. 2022). Cell phenotyping identifies individual cell phenotypes from the measured marker expression values of the cell and directly affects the subsequent cell population analysis results.

The two most common approaches for cell phenotyping in mIF are manual gating and graph-based multivariate clustering. In manual gating, each sample is visualized separately to determine a threshold, and super-threshold cells are labeled as marker positive. This procedure is repeated for all marker channels and slides, and the phenotypes are determined by combining combinations of marker-positive cells (B. Chen et al. 2021; Lin et al. 2023). An example for cell phenotyping from gated markers is presented in Table 2.1. Alternatively, multivariate graph-based clustering is adapted from other single-cell assays. This approach first performs cell clustering, then assigns a phenotype to each cell group based on their average expression profile. Multivariate graph-based clustering is implemented with various modifications across many software packages. Unfortunately, both methods are labor intensive, and their accuracy suffers from image noise and spatial artifacts in mIF images that cause marker expression histograms to appear continuous or uni-modal . As a result, both phenotyping methods possess shortcomings that cannot be ignored. On one hand, manual gating can be subjective.

Table 2.1: The table on the left shows that cells with positive expression of marker CD19 should be classified as B cell, and cell with positive CK expression should be classified as Tumore cell. The table on the right is a hypothetical table of cell marker expression and cell type classification based on it.

(a) Cell type and marker correspondence

Cell Types	Markers
B cells	CD19+
Tumor	CK+

(b) Example of cell phenotyping based on marker expression.

Cell ID	CD19+	CK+	Tumor	B cells
1	TRUE	TRUE	TRUE	TRUE
2	TRUE	FALSE	TRUE	FALSE
3	FALSE	FALSE	FALSE	FALSE

On the other hand, graph-based clustering results are prone to over-clustering and producing poor separation between clusters.

2.2 Previous works

The challenges described above are well recognized and there are a few methods and software developed that attempt to automate cell phenotyping for mIF images. For example, CellSighter is a recently proposed supervised deep-learning algorithm for cell phenotyping that requires a “gold standard” training dataset (Amitay et al. 2023). Another recent solution, ASTIR (Automated assignment of cell identity from single-cell multiplexed imaging and proteomic data), is a fast unsupervised approach that defines cell phenotypes from segmented cell-level data by using a neural network-based mixture model assuming a multivariate log-normal distribution (Geuenich et al. 2021). Instead of binary outputs like in classification methods, ASTIR returns posterior probabilities of different cell types for each cell. This type of output is advantageous because it offers more information than nominal cell types and leaves cell labeling to the clinician’s discretion. Lastly, (Ahmadian et al. 2022) treat the analysis as a pixel classification problem and design a single-step framework for mIF phenotyping that is integrated with other preprocessing steps.

Nevertheless, inconsistencies persist in the results rendered by these learning-based methods when applied across markers, slides, batches, and datasets. These inconsistencies result from the immense variation in the cell-level distribution of phenotyping markers that are often too nuanced to be removed by existing batch correction methods (Wilson et al. 2021; Hunt et al. 2022). For these reasons, it is difficult to fully automate the cell phenotyping process, despite the availability of automated tools, and manual gating is still used to perform cell phenotyping because it is easy to visualize and evaluate the quality of the phenotype.

2.3 Methods

Since automated methods cannot be run without evaluation and supervised methods require a gold-standard dataset, no method is truly fully automated. As a solution, we develop an explicitly semi-automated algorithm called GammaGateR. GammaGateR allows the user to easily perform cell phenotyping, visualize results, and conduct interpretable quality control while reducing manual labor. Based on a novel closed-form Gamma mixture model (cfGMM), GammaGateR is a probabilistic model that is fitted to each channel and slide separately, and outputs positive-component probabilities for each marker. These can then be easily thresholded and combined for semi-automated marker gating or input directly into downstream analysis. GammaGateR has important technical advantages, including 1) improved computation time and model convergence due to its novel closed-form property, and 2) high consistency and reproducibility for phenotyping results across mIF data batches due to incorporation of parameter boundary constraints. In applications on real-world mIF data, we find that GammaGateR has fast and consistent results across many slides and markers. We provide an open-source implementation of our method in the new GammaGateR R package.

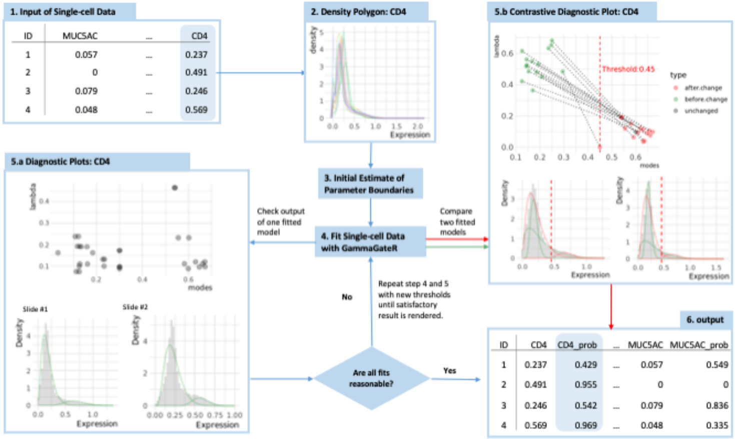
2.3.1 GammaGateR

The GammaGateR algorithm is unique to existing methods for its focus on parsimoniously modeling cell-level marker expression densities. This approach yields tailored-to-slide model estimation in cell-level mIF data where marker expression distributions can vary substantially across slides. The algorithm uses a zero-inflated two-component GMM to model marker expression for each slide. The Gamma mixture model naturally identifies marker-positive and marker-negative cell distributions and returns the probability of belonging to the marker-positive cell distribution for each cell. The returned probabilities can either be used directly in subsequent analysis or combined and dichotomized to define cell phenotypes. GammaGateR incorporates user-specified constraints to provide consistent model fit across a large number of slides. The model evaluation methods included in GammaGateR allow the user to evaluate the constraints and quality check results. The power source of GammaGateR is the closed-form Gamma mixture model, which is a novel approach to phenotyping for mIF data that makes it more computationally efficient than traditional GMMs.

The analysis pipeline is illustrated for the CD4 marker channel (Figure 2.1). GammaGateR takes a single-cell image dataset, with each row corresponding to an individual cell, and each column as the normalized intensity of a marker channel for each cell (Figure 2.1 1). The first step is selecting biological constraints for model fitting by visualizing overlay histograms for each marker channel (Figure 2.1 2). The constraints are not manual thresholds, but represent boundaries for the mode of each component of the fitted distribution across all slides in the dataset. Because marker-positive cells often are a small proportion of all cells and have higher expression values, we limit the mode of the higher component to be no lower than the "elbow" of the overlay histograms (Figure 2.1 2, e.g. 0.45). While GammaGateR can be fit without constraints, the constraints provide more consistent model estimation across many slides. Given the data and constraints, GammaGateR generates the parameter estimates of the Gamma mixture model including modes and proportion of each component, and the posterior and marginal probabilities of each cell being marker-positive.

To ensure accurate model fitting, GammaGateR includes functionality for users to evaluate model fit and modify the fit when needed. Diagnostic plots for the fitted GammaGateR model object consist of a scatter plot of all the slides fitted model modes and lambda (marker-positive probability), and the fitted density curve over the cell expression histogram for each slide in the data set (Figure 2.1 5.a). The x-axis of the scatter plot is the mode for the marker-positive component, and the y-axis is the proportion of the corresponding component. The scatter plot is useful for identifying slides that are outliers with respect to where the mode of marker-positive cells lies or the estimated proportion of marker-positive cells. The histograms can be used for visually evaluating model fit for one slide. A good model fit shows an approximate fit of the smooth density line to the histogram with a marker-positive cell distribution sitting to the right (Figure 2.1 5.a). If there is poor model fit, users can compare fitted models between two different constraints to check how different boundaries affect fitted values (Figure 2.1 5.b). Figure 2.1 5.b compares the model fit for CD4 with no constraints (green) to the model fit with an initial constraint with a lower bound of 0.45 for the marker-positive component (red). The model without constraints places the marker-positive distribution directly over the marker-negative distribution. Users can adjust the parameter boundaries and fit again until satisfactory fittings are rendered. Finally, the output of the fitted models is easily accessible in the GammaGateR model object (Figure 2.1 6). The vignette in the GammaGateR R package provides a guide to fitting the GammaGateR model in the lung cancer dataset from the VectraPolaris dataset available on Bioconductor.

In the extreme tails, the marker-positive probability might not always be higher than that of marker-negative, due to different variances of the two components' Gamma distributions. Therefore, we apply a correction to the posterior probabilities to force them to be monotonic with respect to the marker values. Specifically, after the first crossing of the density curves of the two components, the density curve of the first component will be forced to be non-increasing, and the density curve of the second component will be forced to be non-decreasing. In addition to the posterior probability, GammaGateR also out-


1. Input of Single-cell Data
2. Density Polygon: CD4
density
Expression
3. Initial Estimate of Parameter Boundaries
4. Fit Single-cell Data with GammaGateR
Check output of one fitted model
Compare two fitted models
5.a Diagnostic Plots: CD4
lambda
modes
Slide #1
Slide #2
Density
Expression
5.b Contrastive Diagnostic Plot: CD4
Threshold:0.45
type
after.change
before.change
unchanged
No
Repeat step 4 and 5 with new thresholds until satisfactory result is rendered.
Are all fits reasonable?
Yes
6. output

1. Input of Single-cell Data

ID	MUCSAC	...	CD4
1	0.057	...	0.237
2	0	...	0.491
3	0.079	...	0.246
4	0.048	...	0.569

6. output

ID	CD4	CD4_prob	...	MUC5AC	MUC5AC_prob
1	0.237	0.429	...	0.057	0.549
2	0.491	0.955	...	0	0
3	0.246	0.542	...	0.079	0.836
4	0.569	0.969	...	0.048	0.335

Figure 2.1: Overview of GammaGateR analysis pipeline for the CD4 marker channel. (1) GammaGateR takes segmented cell-level data as input. (2) Density polygons are used to visualize all slide level histograms and select constraints for model fit (3). (4) After model estimation, (5.a) diagnostic plots are used to evaluate the model fit. (5.b) New constraints can be selected and the refitted model can be compared to a previous model. (6) Expression probabilities can be extracted for downstream analysis from the model objects.

puts the marginal probability of the observed marker value for the marker-positive component. The marginal probability is monotonically increasing in the marker intensities and represents the probability that a marker positive cell is less than the given value, x.

2.3.2 cfGMM

For mIF data, we use the GMM to fit cell marker expression values as a weighted sum of different probability distributions that represent unique cell populations (McLachlan, Lee, and Rathnayake 2019). The Gamma distribution is an excellent model for marker values because the domain of the Gamma distribution is strictly positive and it has the flexibility to model the varying skewed densities seen in mIF marker values (Figure 2.1 5.a). However, GMMs are not scalable for mIF image data, because they rely on computationally inefficient numerical methods to obtain the maximum likelihood estimator (MLE). The slow convergence of the MLE for the GMM makes it prohibitive to apply across a large number of channels, slides, and cells. As a solution, we develop a closed-form GMM (cfGMM) estimation procedure based on a recently developed estimator for the Gamma distribution (Ye and Chen 2017). In addition, to improve computational efficiency, the cfGMM has the benefit of allowing prior constraints on model parameters. With the cfGMM in GammaGateR, we enable the flexibility to include a biologically meaningful range for the mode of each component in the Gamma mixture model. This way, users of GammaGateR can restrict estimation to biologically meaningful values.

2.3.2.1 Derivation

We assume the data is a random sample $x_1, \dots, x_n$ from a K component generalized gamma mixture distribution. The density function of X is

$$P(X = x) = \sum_{k=1}^{K} \lambda_k f(x; a_k, b_k, \gamma_k).$$

and the log-likelihood of the dataset is

$$\ell(\mathbf{x}|\mathbf{a},\mathbf{b},\boldsymbol{\lambda}) = \sum_{i=1}^{n} \log \left\{ \sum_{k=1}^{K} \lambda_k f(x_i|a_k, b_k) \right\} \tag{2.1}$$

For each generalized gamma component k, $\lambda_k \in [0,1]$ are the mixture parameters, $\sum_k \lambda_k = 1$; f denotes the generalized gamma density function; a_k, b_k, γ_k are the parameters for the generalized gamma.

Here, we use the expectation maximization (EM) algorithm [dempster_maximum_1977] for parameter estimation. EM algorithm is a standard approach for parameter estimation in mixture models. It introduces the latent multinomial variable $Z_i = (Z_{i1}, \ldots, Z_{iK})$ into the model and maximizes the expected value of the complete data likelihood [dempster_maximum_1977]. The expectation of the complete data likelihood to be maximized for the generalized gamma distribution is

$$\mathbb{E}_Z \ell(x \mid Z) = \sum_{i=1}^{n} \sum_{k=1}^{K} z_{ik} \log f(x_i; a_k, b_k, \gamma_k),$$

where

$$z_{ik} = \mathbb{P}(Z_{ik} = 1 \mid x_i; \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\gamma}) = \frac{f(x_i|a_k, b_k, \gamma_k)}{\sum_{j=1}^{K} f(x_i|a_j, b_j, \gamma_k)}, \tag{2.2}$$

$\mathbf{a} = (a_1, a_2, \ldots, a_K)$, and $\mathbf{b}$, $\boldsymbol{\lambda}$ are similarly defined vectors.

From here, the maximization of the expectation is now analogous to the maximization of generalized gamma distribution for each component of the mixture model.

The expectation of log-likelihood is

$$\mathbb{E}_{z|x}[\log(L(\mathbf{x}|\mathbf{z}))] = \sum_{i=1}^{n} \sum_{k=1}^{K} z_{ik} \log f_k(x_i) \tag{2.3}$$

and

$$f_k(x) = G(a_k, b_k, \gamma_k) = \frac{\lambda_k x^{a_k \gamma_k - 1} exp\{(-x/b_k)^{\gamma_k}\}}{b_k^{a_k \gamma_k} \Gamma(a_k)} \tag{2.4}$$

where $\gamma_k = 1$.

By the two formulas above2.3, **??** The expected joint log-likelihood is

$$\mathbb{E}_{z|x}[\log(L(\mathbf{x}|\mathbf{z}))] = \sum_{k=1}^{K}\sum_{i=1}^{n} z_{ik}\left(\log\gamma_k - a_k\gamma_k\log b_k - log\Gamma(a_k) + (a_k\gamma_k - 1)\log X_i - (\frac{X_i}{b_k})^{\gamma_k}\right) \tag{2.5}$$

The estimators of each of the K terms of the expected joint log-likelihood are derived as follows:

first take derivative of the expression from the above equation (2.5)

$$\sum_{i=1}^{n} z_{ik}\left(\log\gamma_k - a_k\gamma_k\log b_k - \log\Gamma(a_k) + (a_k\gamma_k - 1)\log X_i - \left(\frac{X_i}{b_k}\right)^{\gamma_k}\right)$$

with respect to a_k, b_k, γ_k separately:

$$\frac{\partial\mathbb{E}_{z|x}[\log(L(\mathbf{x}|\mathbf{z}))]}{\partial a_k} = \sum_{i=1}^{n} z_{ik}(-\psi(a_k) - \gamma_k\log b_k + \gamma_k X_i) = 0(1) \tag{2.6}$$

Note that $\psi(x) = \frac{d}{dx}\log\Gamma(x)$ is digamma function.

$$\frac{\partial\mathbb{E}_{z|x}[\log(L(\mathbf{x}|\mathbf{z}))]}{\partial b_k} = \sum_{i=1}^{n}(z_{ik})(-a_k\gamma_k/b_k + \gamma_k X_i^{\gamma_k} b_k^{-\gamma_k - 1}) = 0(2) \tag{2.7}$$

$$\frac{\partial\mathbb{E}_{z|x}[\log(L(\mathbf{x}|\mathbf{z}))]}{\partial \gamma_k} = \sum_{i=1}^{n} z_{ik}\left(\frac{1}{\gamma_k} - a_k\log b_k + a_k\log X_i - \left(\frac{X_i}{b_k}\right)^{\gamma_k}\log\frac{X_i}{b_k}\right) = 0(3) \tag{2.8}$$

Among which, (2)(2.7) can be solved as

$$\hat{b}_k(a_k,\gamma_k) = \left(\frac{\sum_{i=1}^{n} z_{ik}X_i^{\gamma_k}}{a_k\sum_{i=1}^{n} z_{ik}}\right)^{1/\gamma_k} \quad (4) \tag{2.9}$$

Substitute (2.9) (4) into (2.8) (3):

$$\frac{\partial \mathbb{E}_{z|x}[\log(L(\mathbf{x}|\mathbf{z}))]}{\partial \gamma_k}$$

$$= \sum_{i=1}^{n} z_{ik}/\gamma_k + \sum_{i=1}^{n} a_k z_{ik} \log(\frac{X_i}{b_k}) - \sum_{i=1}^{n} z_{ik}(\frac{X_i}{b_k})^{\gamma_k} \log(\frac{X_i}{b_k})$$

$$= \sum_{i=1}^{n} z_{ik}/\gamma_k + \sum_{i=1}^{n} a_k z_{ik} (\log X_i - \log b_k) - b_k^{-\gamma_k} \sum_{i=1}^{n} z_{ik} X_i^{\gamma_k} (\log X_i - \log b_k)$$

$$= \sum_{i=1}^{n} z_{ik}/\gamma_k + \sum_{i=1}^{n} a_k z_{ik} \log X_i - \log b_k \sum_{i=1}^{n} a_k z_{ik} - b_k^{-\gamma_k} \sum_{i=1}^{n} z_{ik} X_i^{\gamma_k} \log X_i + b_k^{-\gamma_k} \log b_k \sum_{i=1}^{n} z_{ik} X_i^{\gamma_k}$$

$$= \sum_{i=1}^{n} z_{ik}/\gamma_k + \sum_{i=1}^{n} a_k z_{ik} \log X_i - \log b_k \sum_{i=1}^{n} a_k z_{ik} - b_k^{-\gamma_k} \sum_{i=1}^{n} z_{ik} X_i^{\gamma_k} \log X_i \frac{a_k \sum_{i=1}^{n} z_{ik}}{\sum_{i=1}^{n} z_{ik} X_i^{\gamma_k}} \log b_k \sum_{i=1}^{n} z_{ik} X_i^{\gamma_k}$$

$$= \sum_{i=1}^{n} z_{ik}/\gamma_k + \sum_{i=1}^{n} a_k z_{ik} \log X_i - \log b_k \sum_{i=1}^{n} a_k z_{ik} - b_k^{-\gamma_k} \sum_{i=1}^{n} z_{ik} X_i^{\gamma_k} \log X_i + a_k \sum_{i=1}^{n} z_{ik} \log b_k$$

$$= \sum_{i=1}^{n} z_{ik}/\gamma_k + a_k \sum_{i=1}^{n} z_{ik} \log X_i - \frac{a_k \sum_{i=1}^{n} z_{ik}}{\sum_{i=1}^{n} z_{ik} X_i^{\gamma_k}} \sum_{i=1}^{n} z_{ik} X_i^{\gamma_k} \log X_i$$

$$= \sum_{i=1}^{n} z_{ik}/\gamma_k + a_k \left(\sum_{i=1}^{n} z_{ik} \log X_i - \frac{\sum_{i=1}^{n} z_{ik}}{\sum_{i=1}^{n} z_{ik} X_i^{\gamma_k}} \sum_{i=1}^{n} z_{ik} X_i^{\gamma_k} \log X_i \right) = 0$$

Solving this, we have

$$\hat{a}_k(\gamma_k) = \frac{\sum_{i=1}^{n} \frac{z_{ik}}{\gamma_k}}{\frac{\sum_{i=1}^{n} z_{ik}}{\sum_{i=1}^{n} z_{ik} X_i^{\gamma_k}} \sum_{i=1}^{n} z_{ik} X_i^{\gamma_k} \log X_i - \sum_{i=1}^{n} z_{ik} \log X_i} (5) \tag{2.10}$$

Plug $\gamma_k = 1$ in (5) (2.10), we now have

$$\begin{aligned}
\hat{a}_k(\gamma_k = 1) &= \frac{\displaystyle\sum_{i=1}^{n} z_{ik}}{\dfrac{\displaystyle\sum_{i=1}^{n} z_{ik}}{\displaystyle\sum_{i=1}^{n} z_{ik} X_i} \displaystyle\sum_{i=1}^{n} z_{ik} X_i \log X_i - \sum_{i=1}^{n} z_{ik} \log X_i} \\
&= \left(\frac{\sum_{i=1}^{n} z_{ik} X_i \log X_i}{\sum_{i=1}^{n} z_{ik} X_i} - \frac{\sum_{i=1}^{n} z_{ik} \log X_i}{\sum_{i=1}^{n} z_{ik}} \right)^{-1} \\
&= \frac{\displaystyle\sum_{i=1}^{n} z_{ik} \sum_{i=1}^{n} z_{ik} X_i}{\displaystyle\sum_{i=1}^{n} z_{ik} \sum_{i=1}^{n} z_{ik} X_i \log X_i - \sum_{i=1}^{n} z_{ik} \log X_i \sum_{i=1}^{n} z_{ik} X_i}
\end{aligned} \tag{2.11}$$

$$\begin{aligned}
\hat{b}_k(\hat{a}_k, \gamma_k = 1) &= \frac{\displaystyle\sum_{i=1}^{n} z_{ik} X_i}{\hat{a}_k \displaystyle\sum_{i=1}^{n} z_{ik}} \\
&= \frac{\displaystyle\sum_{i=1}^{n} z_{ik} \sum_{i=1}^{n} z_{ik} X_i \log X_i - \sum_{i=1}^{n} z_{ik} \log X_i \sum_{i=1}^{n} z_{ik} X_i}{\left(\displaystyle\sum_{i=1}^{n} z_{ik} \right)^2}
\end{aligned} \tag{2.12}$$

In addition, $\hat{\lambda}_k$ can simply be estimated as

$$\hat{\lambda}_k = \frac{\displaystyle\sum_{i=1}^{n} z_{ik}}{n} \tag{2.13}$$

It is worth noting that we are not maximizing the exact Gamma distribution, therefore the algorithm we devise here is an EM-type algorithm.

2.4 Results

2.4.1 Simulation for cfGMM

To compare the bias and compute time of the closed-form GMM to maximum likelihood GMM implementation, we run the cfGMM, the constrained cfGMM, and the GMM to evaluate bias and variance in a sample size of 10,000 across 1,000 simulations. We simulate a two-component mixture model with parameters $\boldsymbol{\lambda} = (0.3, 0.7), \boldsymbol{a} = (0.5, 8), \boldsymbol{b} = (0.5, 1/3)$. For the constrained estimator, we restrict the mode of each component to be in the range $(-\infty, 0)$ and $(0, 5)$ for marker negative and marker positive components, respectively, which include the true mode for each component, 0 (no mode) and 7/3.

Both closed-form estimation procedures have substantially faster computation time than the MLE while maintaining similarly low bias, as in Figure 2.2. The sample size used in the simulation is roughly similar to that of the cell-level mIF image dataset, which further proves that cfGMM brings computation efficiency to our target application. The closed-form GMM, therefore, enables computationally feasible, precise, and flexible model estimation when applied to a large number of channels and slides using GammaGateR. It is also worth noting that the constrained cfGMM converges slightly faster than without constraints. This implies that when using cfGMM, computational cost can be reduced with proper knowledge of biological priors.

2.4.2 Analysis in real dataset

We use three single-cell imaging datasets to evaluate model performance and demonstrate the use of the GammaGateR analysis pipeline: the Colorectal Molecular Atlas Project (Colon MAP) dataset (B. Chen et al. 2021), the Spatial Colorectal Cancer (CRC) Atlas dataset (Heiser et al. 2023) and Ovarian Cancer dataset (Steinhart et al. 2021; J. Wrobel and Ghosh

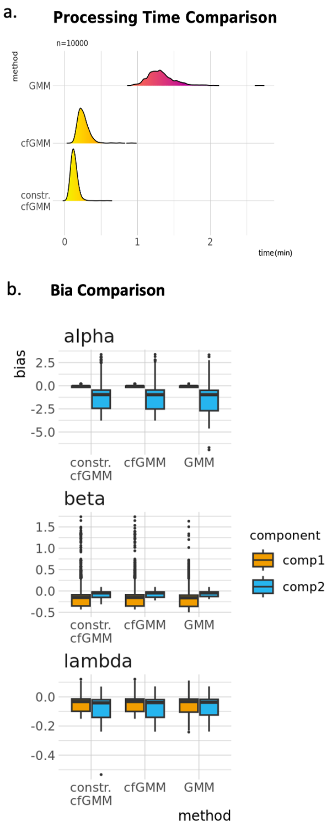


Figure 2.2: Simulation results for cfGMM performance evaluation. a) Run time comparison (in minutes) for three methods: GMM, cfGMM and cfGMM with constraints. b) Estimated bias across 1,000 simulations in a sample size of 10,000.

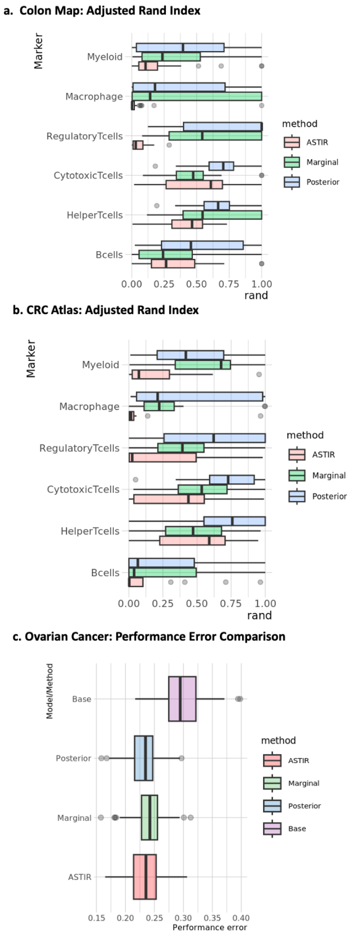


Figure 2.3: Performance evaluation for GammaGateR on the three datasets. “Posterior” and “marginal” refer to the posterior and marginal probabilities from GammaGateR, respectively. Cell phenotyping performance comparing GammaGateR to ASTIR in the (a) Colon MAP and (b) CRC Spatial atlas. (c) Survival prediction performance error in the ovarian cancer dataset, measured by 1-C-index. “Base” indicates

2022). After processing and prior to analysis, cell expression values were normalized by first mean division then log10 transformation to reduce slide-to-slide variation (Harris et al. 2022).

To compare the methods in determining cell phenotypes we assess the accuracy of each method relative to a “silver standard” manual phenotyping in the Colon MAP and CRC atlas datasets and evaluate the efficacy in predicting survival in the ovarian cancer data. We compare phenotyping results obtained using GammaGateR and ASTIR to “silver standard” manual phenotyping using the Adjusted Rand Index (Hubert and Arabie 1985). Adjusted rand index typically takes values between 0 and 1, where a larger value indicates a better alignment between two categorical variables. The silver standard is obtained by gating the raw images based on visual inspection and defining marker-positive cells as those that contain more than 50% marker-positive pixels within each cell. Semi-automated marker gating results are obtained using GammaGateR as described in Section 2.3.1, with monotonically adjusted posterior probability and marginal probability thresholded at 0.5 to define marker positive cells. The same phenotype definitions used for ASTIR are used to define phenotypes from marker positive labels using GammaGateR and manual marker gating as well. Each cell belongs to a given phenotype if it is marker positive for combinations of markers for that phenotype. ASTIR phenotypes are determined by selecting the cell type with the maximum probability for each cell. All methods use the same combinations of markers to define phenotypes. For both datasets and all cell types in these datasets, the posterior probability by GammaGateR yields higher Median ARI (Figure 2.3 a & b). This means that the posterior probability has consistently greater similarity to the silver standard than marginal probability and ASTIR (Figure 2.3 a & b). However, for some cell types (e.g. Macrophage, B-cells, Myeloid), all methods have low performance. This is an indication of systematic difference in how the algorithms identify positive cells relative to the manual labels.

Because the Ovarian Cancer dataset does not include manual cell phenotypes, we instead compare the prediction accuracy of survival time data for each method across all patients in the study to determine if one method has greater biological sensitivity than the other. The original study shows that survival of ovarian cancer patients is significantly correlated with B-cell and CD4 T-cell, as well as spatial interaction between CD4 T-cell and macrophage. Therefore, we evaluate the methods using this dataset by fitting a survival model with age, cancer stage, B-cell proportions, CD4 T-cell proportions, and the spatial interaction between Macrophage and CD4 T-cells estimated by Ripley's H with r=50 (Kiskowski, Hancock, and Kenworthy 2009). Ripley's H is a geospatial index that describes spatial attraction/repulsion. We fit a model for each method, where the cell phenotypes are determined using the given method, and compare all models, as well as a base model that includes only age and cancer stage. We use a random forest survival model to be sensitive to complex nonlinear relationships (Ishwaran et al. 2008; Ishwaran, Kogalur, and Kogalur 2023). To estimate variability in the out-of-bag performance error, we compare the methods across 100 bootstrap samples. Performance error quantification is based on C-index (Harrell et al. 1982), where a low performance error means that the model is a good fit, and a performance error of 0.5 is a random chance. For all methods, incorporating the cell-level information reduces out-of-bag error performance by approximately 0.075, over the base model that includes only age and cancer stage. This indicates that spatial cell phenotype covariates are useful in predicting survival outcomes, consistent with the original findings in Steinhart et al. (2021). The posterior probability slightly outperforms other methods, having the lowest prediction error in 46% of the bootstrap samples, compared to 36% with ASTIR and 18% with the marginal probability (Figure 2.3 c).

2.5 Summary

GammaGateR is a semi-automated maker gating tool. Driven by a novel cfGMM estimation framework, GammaGateR generates reproducible and evaluable marker gating for mIF data. In addition, cfGMM enables computationally feasible model estimation for large-scale datasets like mIF. The marker gating output of GammaGateR can be used to define phenotypes and as input to downstream analysis. GammaGateR implements interactive visualization to quality check the clustering results (see vignette in Supplementary Material) and allows users to modify the constraints to improve model results when needed. Consequently, GammaGateR provides more consistent results with silver standard labels than ASTIR, the existing state-of-the-art method for automated phenotyping of cell-level mIF data. In the examples shown in this section, GammaGateR phenotypes had slightly improved ovarian cancer survival prediction accuracy compared to ASTIR. This paper also compares the posterior and the marginal probabilities returned by GammaGateR. The marginal probabilities only use the marker positive cell distribution to determine cell phenotypes, whereas the posterior probabilities take into account the distribution of the marker negative cells. Using posterior probabilities was almost always better indicating the importance of accounting for the full distribution of the marker intensities when identifying marker-positive cells.

3 Chapter 3: Missing data imputation in mIF imaging

Just like all data created and collected by human beings, missing data is inevitable in mIF image as well. Bao et al. (2021) in their paper gave a brief summary of types of missing data in mIF image, as in Figure 3.1. Case 1 in Figure 3.1 refers to the missingness of one or more entire marker channel due to low image quality, which occurs roughly 3% of all times. Other possible reasons for missing channel, not described in Bao et al. (2021), can be supply shortage in certain type of fluorescent material or change in research plan. Interestingly, all the existing marker channel applications are inspired by the expensiveness of high-plex image instead. Due to time and financial constraints, mIF with no more than seven channels are often more feasible to obtain than the 40-channel mIFs(Wu et al. 2023). To break the restraint in cell phenotypes availability from few number of markers, imputation of marker channels are proposed. Case 2 Figure 3.1 occurs more frequently, when tissue wears off in the cycles of staining - wash off described in Figure 1.2.

Owing to the rapid development in the field of computer vision, all current applications in mIF imputation are implemented with machine learning and/or deep learning methods. In the three applications covered in this document today, Bao et al. (2021) use generative adversarial networks (GANs), Wu et al. (2023) use gradient boosting decision tree in combination with convolutional neural network, and Sims and Chang (2023) use masked autoencoders (MAE). All applications performs well in the evaluations set up in the paper. However, the subsequent analysis with imputed image can benefit from statistical thinking in data imputation. This will be discussed further in chapter 4.

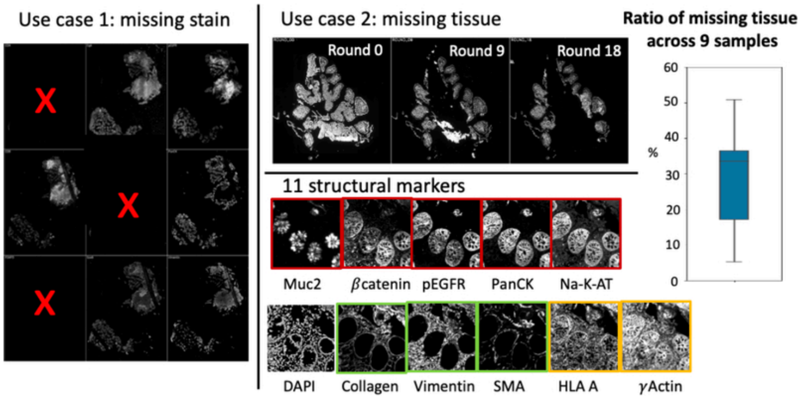


Figure 3.1: Types of missing data in mIF. Image courtesy of Bao et al. (2021).

3.1 Application case 1: Missing tissue imputation

3.1.1 Method: GANs

The fundamental version of GANs comprises of two compartments: a discriminator and a generator (Goodfellow et al. 2014). Figure 3.2 by Bok and Langr (2019) gives a brief sketch of how GANs works. Like a turn-based strategy game, the two components take turns to run an epoch. Starting with a noise distribution (usually a uniform distribution), the generator's goal is to generate data that is close to the real data. The discriminator's goal is to identify the real data between a mix of real data and data generated by the generator. With classification error fed back to generator and discriminator, both opponents update their weights: the generator will try to maximize the probability that the discriminator misclassify generated data as real, and the discriminator will try to maximize classification accuracy. Within infinite number of rounds, they will eventually reach a state close to equilibrium, where either party can only improve negligibly: generator generates close-to-real data, and discriminator classifies with 50% accuracy(Bok and Langr 2019). This is the point where the algorithm stops.

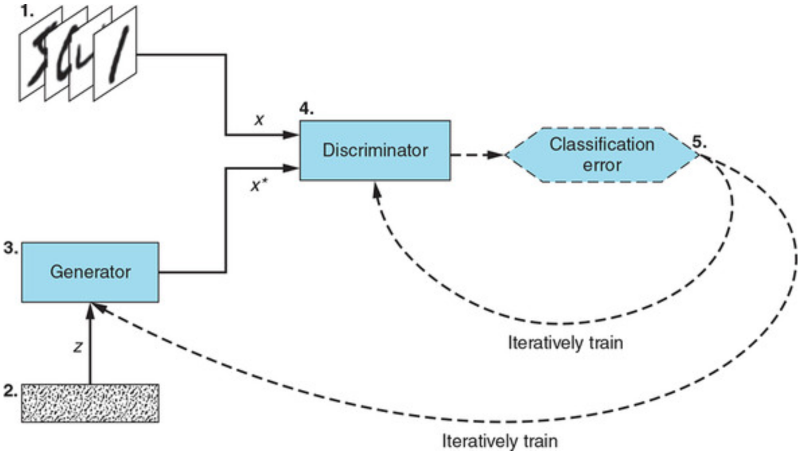


Figure 3.2: Illustration of how GANs work. 2. represents noise data input, and x^* is the "fake" data generated by 3. generator. The goal of the generator is to produce data very similar to real data. 1. represents real data (some hand written digits. Real data x is sent to 4. discriminator along with x^*. The discriminator then try to distinguish true data from generated data. Finally, 5. the classification error is sent back to both 3 and 4 to iteratively train for better performance. Image courtesy of Bok and Langr (2019)

The value function of GANs can be written as formula as follows. $D(x)$ represents the probability that discriminator classify x as real data. z is the noise distribution. $G(z)$ is the fake data that generator creates based on the noise distribution.

$$\min_G \max_D V(D,G) = E_{x \sim p_{data}(x)}[\log D(x)] + E_{z \sim p_z(z)}[\log(1 - D(G(z)))]$$

Discriminator is trained to maximize $V(D,G)$, and the generator is trained to minimize $E_{z \sim p_z(z)}[\log(1 - D(G(z)))]$, the probability that fake data is recognized and classified as fake data.

One disadvantage of the original GANs is its weak control on the generated data, due to the random noise input. This disadvantage stands out especially with image synthesis. Conditional GANs (CGANs) provided a promising solution to this issue by including auxiliary information on both generator and discriminator (Mirza and Osindero 2014). The auxiliary information is usually data from the same class, for example other images in the case of image synthesis. Suppose x belongs to the input data class, and y is the intended output data class. GANs would learn the mapping of $G : z \rightarrow y$, while CGANs learns $G : x, z \rightarrow y$ (Isola et al. 2017; Souza et al. 2023). The updated value function for CGANs is as follows:

$$\min_G \max_D V(D,G) = E_{x,y}[\log D(x)] + E_{x,z}[\log(1 - D(x, G(x,z)))] \tag{3.1}$$

Based on this, *pix2pix* is developed. It is able to perform image-to-image translation by using image pairs (x, y) to train the data. Figure 3.3 shows the types of (x, y) it use. Intuitively, image imputation can utilize this idea: with sets of channels that does not have missing data, well-trained *pix2pix* can generate imputed data.

3.1.2 Application in mIF: pixN2N-HD

pixN2N-HD is a “novel multi-channel high-resolution image synthesis approach”, an extension based on *pix2pix*. “N2N”

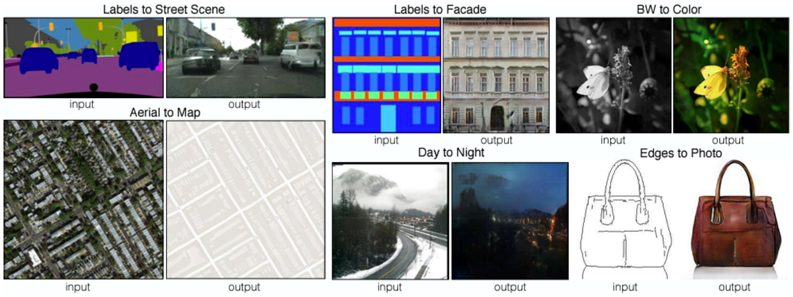


Figure 3.3: Input and output image pair example for pix2pix. Image courtesy of Isola et al. (2017).

represents "N-to-N", which distinguishes itself from the widely-used (N-1)-to-1 model. N represents the number of marker channels, and in the dataset used in this paper, N=11. In (N-1)-to-1 design, 10 channels are used as input and 1 channel is used as output, and this repeats for 11 permutations of models. The "N-to-N" instead uses a random gate strategy, as shown in Figure 3.4. Let δ be the binary vector indicating channels without missing tissue. When $\delta^{(i)}$ is turned on, image of the *ith* channel will go into the generator; when turned off, a blank image will be feed into the generator instead. The value function is similar to Equation 3.1. $X = \delta(M)$, the non-missing channels is the input, and $Y = \bar{\delta}(M)$ is the output.

$$\min_{G}\max_{D} V(D,G) = E_{X,Y}[\log D(X)] + E_{X}[\log(1 - D(X, G(X)))]$$

This paper evaluated the model performance by comparing "N-to-N" model with "(N-1)-to-1" model and another "(N-1)-to-1 random gate" model, which blends in random gate but still needs to train 11 separate models. An index for measuring image similarity, the structure similarity index measure (SSIM) is used to assess whether "N-to-N" model generates comparable results with the other two methods (Wang et al. 2004). The result shows that all pairs of methods do not have significantly different results on a 0.05 significance level, and therefore the methods are concluded to be comparable. This "N-to-N" model take significantly less amount of time to train compared to the

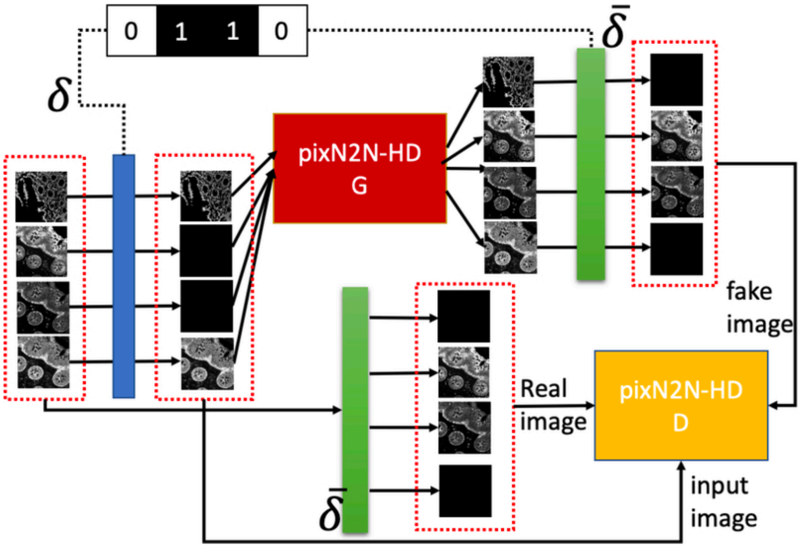


Figure 3.4: Work flow of pixN2N-HD. The topleft binary vector is the random gate, where 1 represents channels with missing tissue and 0 represents channels without missing tissue. Images from intact marker channels are sent to the both generator and discriminator as auxiliary information. Generator synthesizes images of all marker channels, and only the image for marker channel that has missing tissues are sent to discriminator. The discriminator then try to classify real and imputed image for the missing channel. Image courtesy of Bao et al. (2021).

other methods, which is very meaningful in terms of effective computation.

3.2 Application case 2: Marker channel imputation

Both 7-UP and CyCIF panel reduction are intended for marker channel imputation, providing access to otherwise expensive high-plex (40+ channels) mIF image for study that can only obtain low-plex images. Interestingly, the two application uses very different methods for imputation.

3.2.1 Application 2.1: 7-UP

7-UP starts from a 7-plex mIF image and generates high-plex image that can identify up to 16 different cell types (Wu et al. 2023). This approach consists of three main parts:

1. Marker panel selection. This part will select the seven markers to start with, using concrete autoencoder. Concrete autoencoder is an feature selection method, of which the loss function is the difference between the original sample and the reconstructed low-dimension sample (Balın, Abid, and Zou 2019).
2. Morphology feature extraction. This step uses a convolutional neural network to learn the morphology features, i.e. spatial and structural features of cells. Convolutional neural networks are similar to layers of linear regressions, where there are more combinations of weights linked to each input variable.
3. Marker expression imputation. Once the location and structure of cells are learned, the important task left is to impute the expression of each marker on each cell. The imputation is performed using XGBoost, a scalable gradient-boosting tree software (T. Chen and Guestrin 2016).

A series of evaluation and analysis are performed to show the validity of the method. The performance of the method is examined in three ways:

1. Calculating the pearson correlation coefficient between the imputed marker expression and the testing data marker expression.
2. Calculating the F1 score between the imputed and testing data cell type. F1-score is the harmonic mean of precision and sensitivity: $2/(sensitivity^{-1}+precision^{-1})$. Cell type is generated from the marker expression through k-nearest neighbor.
3. Patient survival status, HPV status and disease recurrence are used to further evaluate the cell type outcomes. AUC score for patient status prediction is calculated for both imputed data outcome and training data.

All evluation shows that the imputation generates comparable results with the training data, hence proven the validity of this method.

3.2.2 Application 2.2: CyCIF panel reduction

This method is intended to be an improvement from their own previous work (Ternes et al. 2022). The previous work first go through panel selection and then imputes marker channel with variatioal autoencoder. The current improved method (Sims and Chang 2023) uses masked autoencoder for image synthesis as shown is Figure 3.5. The difference is the adoption of within-model iterative selection of marker panels, as the authors believe that panel selection should be more closely tied with panel reconstruction. Starting with standard DAPI, each marker is added to the panel, predict marker intensities of other panels, and mean Spearman correlation is calculated between the predicted intensity and real intensity. The marker with highest correlation is selected, and the next round continues until the panel is constructed. The ratio of masked channels depends on tasks, though 25%~75% is a reasonable range.

The method outcome is evaluated by Spearman correlation with the true data. It is shown in the results that both MAE and the iterative panel selection outperforms the VAE and out-of-box panel selection of the previous method.

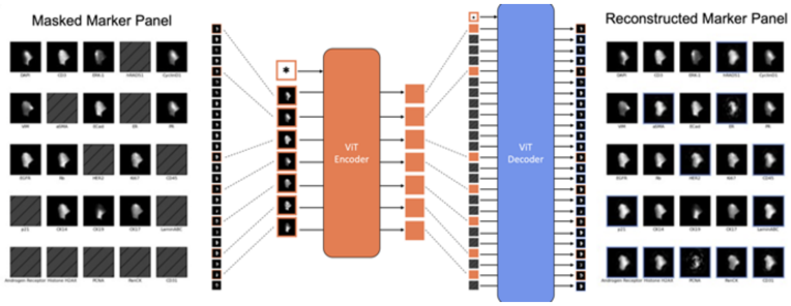


Figure 3.5: CyCIF panel reduction with autoencoder. The left hand side is the masked marker panel, where the mask represents missing marker channels. Next, in the encoder step, the masked markers are translated to a lower-dimnesion latent space. The masked panels are then "imputed" through the decoder step, where latent space information are tranlated back to the same dimension space as the imput image, creating full panel of desired markers. Figure courtesy of Sims and Chang (2023).

4 Chapter 4: Future directions

In the missing data scenarios described in Chapter 3, which channel or tisue will be missing cannot be predicted. These can be seen as missing completely at random (MCAR), i.e. the reason for missing is irrelevant with how the image looks like. However, for the low-plex image channel imputation methods, the markers are selected by machine learning methods to maximize the accuracy. These however, are missing not at random (MNAR), i.e. the reason for "missing" (channels to be imputed) are correlated with the image itself. For this reason, this type of imputation results should be taken with a grain of salt. It might be possible to deduce the amount of bias this creates through statistical inference, but for now this is not in the scope of this project. The following two considerations are ignored in the works mentioned in Chapter 3. In the future, we can perform an evaluation and see how much the imputation results improve when the following statistical properties are considered.

4.1 Multiple imputation

All methods in Chapter 3 evaluated the accuracy of their result by either comparing with the imputation outcome of a previous method or with a held-out evaluation dataset. In addition, Wu et al. (2023) in Section 3.2.1 used the imputation output for cell phenotyping and predicting patient phenotypic outcomes. In principle, this is not the best practice of subsequent analysis with the imputation outcome. Point estimates from such imputation underestimates the standard deviation of the point estimates, as shown below.

Rubin (1996) states that it is important to be statistically valid for estimates of scientific estimands, such as population mean.

Statistical validity for an estimand means an at least approximately non-biased point estimate, and an statistical test that rejects the null hypothesis no more than 5% of the time, when the nominal significance level is 5% (Rubin 1996; Van Buuren 2018). Under such guideline, we can check the multiple imputation estimand's expectation and variance. Van Buuren (2018) presented a clear interpretation of such, conditioning on observed data: Let Y_{mis} be missing data, and Y_{obs} be observed data, Q be the population value and $\hat{Q}$ is the estimate of Q. The posterior distribution of Q given observed data is

$$P(Q|Y_{obs}) = \int P(Q|Y_{obs}, Y_{mis})P(Y_{mis}|Y_{obs})dY_{mis}$$

The posterior mean of Q is therefore

$$E[Q|Y_{obs}] = E[E(Q|Y_{obs}, Y_{mis})|Y_{obs}]$$

Which can be interpreted as the average of estimates over repeatedly imputed data, given the observed data.

The posterior variance is therefore

$$Var(Q|Y_{obs}) = E[Var(Q|Y_{obs}, Y_{mis})|Y_{obs}] + Var[E(Q|Y_{obs}, Y_{mis})|Y_{obs}]$$

Which can be interpreted as the within-variance of the estimate in each imputed data, and the between-variance among the repeatedly imputed data.

In this case, single imputation would be an unbiased estimator. However, it will be biased in variance estimation, as it does not incorporate between-variance at all (the second proportion of the last equation RHS). Therefore, for better accuracy with estimate variance, multiple imputation should be used. The subsequent project could use the colon map data in GammaGateR project, where some channels are missing, and compare the validity of confidence interval for the estimated quantities.

4.2 Imputation with patient information

As described in Wu et al. (2023), cell phenotypes are usually considered with certain prognostic information, such as patient survival, recurrence, and disease status. That is, suppose we want to estimate a certain parameter θ, patient parameters X, we are assuming

$$\theta = E[g(X, Y_{obs}, Y_{mis})]$$

and hence

$$Y_{mis} = f(X, Y_{obs}) + \epsilon, \epsilon \perp (X, Y_{obs})$$

The existing imputation methods all uses $Y_{mis} = f(Y_{obs}) + \epsilon$. That is, there exists bias of

$$f(X, Y_{obs}) - f(Y_{obs})$$

This bias can only be ignored if we assume that $Y_{mis} \perp X | Y_{obs}$. That is, patient information does not contribute to the prediction of missing data given observed data. This is a bold assumption to make, and should be avoided unless with concrete supporting prior knowledge.

References

Ahmadian, Mansooreh, Christian Rickert, Angela Minic, Julia Wrobel, Benjamin G Bitler, Fuyong Zing, Michael Angelo, Elena W Hsieh, Debashis Ghosh, and Kimberly R Jordan. 2022. "A Platform-Independent Framework for Phenotyping of Multiplex Tissue Imaging Data." *bioRxiv*, 2022–10.

Amitay, Yael, Yuval Bussi, Ben Feinstein, Shai Bagon, Idan Milo, and Leeat Keren. 2023. "CellSighter: A Neural Network to Classify Cells in Highly Multiplexed Images." *Nature Communications* 14 (1): 4302.

Balın, Muhammed Fatih, Abubakar Abid, and James Zou. 2019. "Concrete Autoencoders: Differentiable Feature Selection and Reconstruction." In *International Conference on Machine Learning*, 444–53. PMLR.

Bao, Shunxing, Yucheng Tang, Ho Hin Lee, Riqiang Gao, Sophie Chiron, Ilwoo Lyu, Lori A Coburn, et al. 2021. "Random Multi-Channel Image Synthesis for Multiplexed Immunofluorescence Imaging." In *MICCAI Workshop on Computational Pathology*, 36–46. PMLR.

Bok, Vladimir, and Jakub Langr. 2019. *GANs in Action: Deep Learning with Generative Adversarial Networks*. Simon; Schuster.

Chen, Bob, Scurrah Cherie'R, Eliot T McKinley, Alan J Simmons, Marisol A Ramirez-Solano, Xiangzhu Zhu, Nicholas O Markham, et al. 2021. "Differential Pre-Malignant Programs and Microenvironment Chart Distinct Paths to Malignancy in Human Colorectal Polyps." *Cell* 184 (26): 6262–80.

Chen, Tianqi, and Carlos Guestrin. 2016. "Xgboost: A Scalable Tree Boosting System." In *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 785–94.

Coons, Albert H, Hugh J Creech, and R Norman Jones. 1941. "Immunological Properties of an Antibody Containing a Fluorescent Group." *Proceedings of the Society for Experimental Biology and Medicine* 47 (2): 200–202.

Duraiyan, Jeyapradha, Rajeshwar Govindarajan, Karunakaran Kaliyappan, and Murugesan Palanisamy. 2012. "Applications of Immunohistochemistry." *Journal of Pharmacy & Bioallied Sciences* 4 (Suppl 2): S307.

Eng, Jennifer, Elmar Bucher, Zhi Hu, Ting Zheng, Summer L Gibbs, Koei Chin, and Joe W Gray. 2022. "A Framework for Multiplex Imaging Optimization and Reproducible Analysis." *Communications Biology* 5 (1): 438.

Geuenich, Michael J, Jinyu Hou, Sunyun Lee, Shanza Ayub, Hartland W Jackson, and Kieran R Campbell. 2021. "Automated Assignment of Cell Identity from Single-Cell Multiplexed Imaging and Proteomic Data." *Cell Systems* 12 (12): 1173–86.

Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. "Generative Adversarial Nets." *Advances in Neural Information Processing Systems* 27.

Graf, John, Sanghee Cho, Elizabeth McDonough, Alex Corwin, Anup Sood, Andreas Lindner, Manuela Salvucci, et al. 2022. "FLINO: A New Method for Immunofluorescence Bioimage Normalization." *Bioinformatics* 38 (2): 520–26.

Harrell, Frank E, Robert M Califf, David B Pryor, Kerry L Lee, and Robert A Rosati. 1982. "Evaluating the Yield of Medical Tests." *Jama* 247 (18): 2543–46.

Harris, Coleman R, Eliot T McKinley, Joseph T Roland, Qi Liu, Martha J Shrubsole, Ken S Lau, Robert J Coffey, Julia Wrobel, and Simon N Vandekar. 2022. "Quantifying and Correcting Slide-to-Slide Variation in Multiplexed Immunofluorescence Images." *Bioinformatics* 38 (6): 1700–1707.

Heiser, Cody N, Alan J Simmons, Frank Revetta, Eliot T McKinley, Marisol A Ramirez-Solano, Jiawei Wang, Justin Shao, et al. 2023. "Molecular Cartography Uncovers Evolutionary and Microenvironmental Dynamics in Sporadic Colorectal Tumors." *bioRxiv*, 2023–03.

Hubert, Lawrence, and Phipps Arabie. 1985. "Comparing Partitions." *Journal of Classification* 2: 193–218.

Hunt, Gregory J, Mark A Dane, James E Korkola, Laura M Heiser, and Johann A Gagnon-Bartsch. 2022. "Systematic Replication Enables Normalization of High-Throughput Imaging Assays." *Bioinformatics* 38 (21): 4934–40.

Hussaini, Haizal Mohd, Benedict Seo, and Alison M Rich. 2022. "Immunohistochemistry and Immunofluorescence." In *Oral Biology: Molecular Techniques and Applications*, 439–50. Springer.

Ishwaran, Hemant, Udaya B Kogalur, Eugene H Blackstone, and Michael S Lauer. 2008. "Random Survival Forests."

Ishwaran, Hemant, Udaya B Kogalur, and Maintainer Udaya B Kogalur. 2023. "Package 'randomForestSRC'." *Breast* 6 (1): 854.

Isola, Phillip, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. "Image-to-Image Translation with Conditional Adversarial Networks." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1125–34.

Kiskowski, Maria A, John F Hancock, and Anne K Kenworthy. 2009. "On the Use of Ripley's k-Function and Its Derivatives to Analyze Domain Size." *Biophysical Journal* 97 (4): 1095–1103.

Lin, Jia-Ren, Shu Wang, Shannon Coy, Yu-An Chen, Clarence Yapp, Madison Tyler, Maulik K Nariya, et al. 2023. "Multiplexed 3D Atlas of State Transitions and Immune Interaction in Colorectal Cancer." *Cell* 186 (2): 363–81.

McKinley, Eliot T, Justin Shao, Samuel T Ellis, Cody N Heiser, Joseph T Roland, Mary C Macedonia, Paige N Vega, Susie Shin, Robert J Coffey, and Ken S Lau. 2022. "MIRIAM: A Machine and Deep Learning Single-Cell Segmentation and Quantification Pipeline for Multi-Dimensional Tissue Images." *Cytometry Part A* 101 (6): 521–28.

McLachlan, Geoffrey J, Sharon X Lee, and Suren I Rathnayake. 2019. "Finite Mixture Models." *Annual Review of Statistics and Its Application* 6: 355–78.

Mirza, Mehdi, and Simon Osindero. 2014. "Conditional Generative Adversarial Nets." *arXiv Preprint arXiv:1411.1784*.

Ramos-Vara, Jose A. 2005. "Technical Aspects of Immunohistochemistry." *Veterinary Pathology* 42 (4): 405–26.

Rubin, Donald B. 1996. "Multiple Imputation After 18+ Years." *Journal of the American Statistical Association* 91 (434): 473–89.

Schüffler, Peter J, Denis Schapiro, Charlotte Giesen, Hao AO Wang, Bernd Bodenmiller, and Joachim M Buhmann. 2015.

"Automatic Single Cell Segmentation on Highly Multiplexed Tissue Images." *Cytometry Part A* 87 (10): 936–42.

Schürch, Christian M, Salil S Bhate, Graham L Barlow, Darci J Phillips, Luca Noti, Inti Zlobec, Pauline Chu, et al. 2020. "Coordinated Cellular Neighborhoods Orchestrate Antitumoral Immunity at the Colorectal Cancer Invasive Front." *Cell* 182 (5): 1341–59.

Sheng, Wenjie, Chaoyu Zhang, TM Mohiuddin, Marwah Al-Rawe, Felix Zeppernick, Franco H Falcone, Ivo Meinhold-Heerlein, and Ahmad Fawzi Hussain. 2023. "Multiplex Immunofluorescence: A Powerful Tool in Cancer Immunotherapy." *International Journal of Molecular Sciences* 24 (4): 3086.

Sims, Zachary, and Young Hwan Chang. 2023. "A Masked Image Modeling Approach to Cyclic Immunofluorescence (CyCIF) Panel Reduction and Marker Imputation." *bioRxiv*, 2023–05.

Souza, Vinicius Luis Trevisan de, Bruno Augusto Dorta Marques, Harlen Costa Batagelo, and João Paulo Gois. 2023. "A Review on Generative Adversarial Networks for Image Generation." *Computers & Graphics*.

Steinhart, Benjamin, Kimberly R Jordan, Jaidev Bapat, Miriam D Post, Lindsay W Brubaker, Benjamin G Bitler, and Julia Wrobel. 2021. "The Spatial Context of Tumor-Infiltrating Immune Cells Associates with Improved Ovarian Cancer Survival." *Molecular Cancer Research* 19 (12): 1973–79.

Ternes, Luke, Jia-Ren Lin, Yu-An Chen, Joe W Gray, and Young Hwan Chang. 2022. "Computational Multiplex Panel Reduction to Maximize Information Retention in Breast Cancer Tissue Microarrays." *PLoS Computational Biology* 18 (9): e1010505.

Van Buuren, Stef. 2018. *Flexible Imputation of Missing Data*. CRC press.

Wang, Zhou, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. 2004. "Image Quality Assessment: From Error Visibility to Structural Similarity." *IEEE Transactions on Image Processing* 13 (4): 600–612. https://doi.org/10.1109/TIP.2003.819861.

Wilson, Christopher M, Oscar E Ospina, Mary K Townsend, Jonathan Nguyen, Carlos Moran Segura, Joellen M Schildkraut, Shelley S Tworoger, Lauren C Peres, and Brooke L Fridley. 2021. "Challenges and Opportunities in the Statistical Analysis of Multiplex Immunofluorescence Data." *Cancers* 13 (12): 3031.

Wrobel, J, and T Ghosh. 2022. "VectraPolarisData: Vectra Polaris and Vectra 3 Multiplex Single-Cell Imaging Data." *Bioconductor R Package Version* 1.

Wrobel, Julia, Coleman Harris, and Simon Vandekar. 2023. "Statistical Analysis of Multiplex Immunofluorescence and Immunohistochemistry Imaging Data." In *Statistical Genomics*, 141–68. Springer.

Wu, Eric, Alexandro E Trevino, Zhenqin Wu, Kyle Swanson, Honesty J Kim, H Blaize D'Angio, Ryan Preska, et al. 2023. "7-UP: Generating in Silico CODEX from a Small Set of Immunofluorescence Markers." *PNAS Nexus* 2 (6): pgad171.

Ye, Zhi-Sheng, and Nan Chen. 2017. "Closed-Form Estimators for the Gamma Distribution Derived from Likelihood Equations." *The American Statistician* 71 (2): 177–81.