



HAL
open science

Artificial Intelligence and generative data augmentation for automated chromosomal aberration detection in cytogenetic FISH imaging

Quentin Tallon, Emmanuel Moebel, Juan Martinez, Eric Gregoire, Pascale Fernandez, Delphine Dugue, Rose Deboucha, Laure Bobyk, Marco Valente, Gaetan Gruel, et al.

► To cite this version:

Quentin Tallon, Emmanuel Moebel, Juan Martinez, Eric Gregoire, Pascale Fernandez, et al.. Artificial Intelligence and generative data augmentation for automated chromosomal aberration detection in cytogenetic FISH imaging. 48h European Radiation Research Society Meeting (ERRS), Sep 2024, Aveiro, Portugal. hal-04902922

HAL Id: hal-04902922

<https://inria.hal.science/hal-04902922v1>

Submitted on 21 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

ARTIFICIAL INTELLIGENCE AND GENERATIVE DATA AUGMENTATION FOR AUTOMATED CHROMOSOMAL ABERRATION DETECTION IN CYTOGENETIC FISH IMAGING

Quentin TALLON^{1,2}

with E. Moebel², J. Martinez¹, E. Gregoire¹,
P. Fernandez¹, D. Dugue¹, R. Deboucha¹,
L. Bobyk³, M. Valente³, G. Gruel¹,
C. Kervrann², and M.A. Benadjaoud¹

¹ LRAcc, PSE-SANTE/SERAMED IRSN

² SAIRPICO, INRIA Rennes

³ IRBA

Inria

Contexte

Données

Métaphase

Chromosome

Génératif

Biological Dosimetry Context



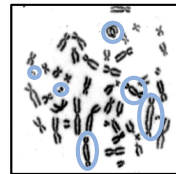
Accidental or malicious exposure of people to ionizing radiation



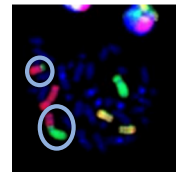
Blood collection and placement in cell culture (48h)



Acquisition of metaphases



GIEMSA

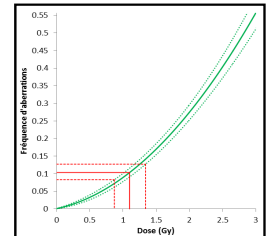


FISH

Counting aberrations (most time-consuming step)

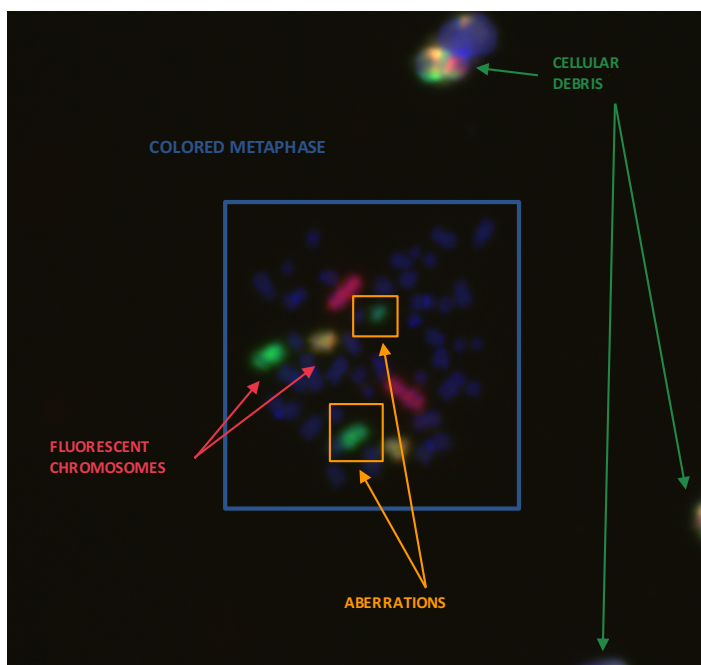
Limitations

- Manual counting
- Cognitive fatigue
- Not enough qualified people
- Complex task: noise, change of acquisition equipment, chromosome polymorphism



Inverse regression of the calibration curve for dose estimation and its uncertainties

FISH Metaphase Imagery



Facilitate the **counting** of chromosomal aberrations by biologists

Two tasks:

- **Localization** of chromosomes of interest
- **Classification** of localized chromosomes

We focus on **fluorescent** chromosome detection

Aberrations visible thanks to **color co-localization**

AI in computer vision through deep learning: need of **images** and **corresponding annotations**

Contexte

Données

Métaphase

Chromosome

Génératif

Aberration Counting Sheets

Manuscript

LAME: 1210 T12 L18		SONDES UTILISEES: 2, 4, 12	Nb de cellules normales: 347 - 4 = 343		
DATE DE L'EXPERIENCE: 18-10-12		OPERATEUR: EG	Nb de cellules Anormales non EC: 2 + 4 = 397		
DATE DE LECTURE: 22-10-12		PAGE N°: 1	Nb de cellules Anormales type EC: 2		
N° métaphase	X	Y	Nb spots	Dessin	Aberrations
Anomale			2 4 12		
			2 4 12		
1	10		5 2 2		
2	31		2 5 2		TR(4;D)
3	64		2 4 5		
4	250		5 2 2		TR(2;D)

Digitized

Dose:	No. Lame: 12-10712L18-A	Chromosomes Projet	Chromosomes Vert	Chromosomes Jaune	Compteur:	Vérificateur:	Distribution total								
							D88 ends	0	0	0	0	0	1	0	0
Générateur Sheets															
Nomenclature Sels	D88 ends/ cellule	Cell. No.	Normale (N)/ Secondaire (S)	Complexe (T/N)	Stable (S/N)	Aberration 1	Aberration 2	Aberration 3	Aberration 4	Aberration 5	Aberration 6	Aberration 7	Aberration 8	Aberration 9	
(2)(2)(X)(Y)(2)	5	30			0	F(2)	Tw1 (2,5)								
(4)(3)	2	31			1	Out (4,8)									
(4)(2)(X)(Y)(2)(X)(Y)(2)(X)(Y)(2)(X)(Y)(2)(X)(Y)(2)	25	30			0	Insertion	Out (10,2)	Insertion	Insertion	Tw1 (4,5)	F(3,2)	Tw1 (12,8)	F(4)	Deletion	
(2)(2)(X)(Y)(2)	4	25			3	Tw1 (4,5)									

- The LRAcc annotation is a **count annotation**:
 - one counting sheet corresponds to one microscope slide
 - one line corresponds to one metaphase
 - we don't know the **spatial position** of aberrations

- Work on **digitization** of handwritten counting sheets: Juan Martinez, Eric Grégoire and Rose Déboucha (LRAcc)

Contexte

Données

Métaphase

Chromosome

Génératif

Need for a spatial annotation: encoding the analysis methodology

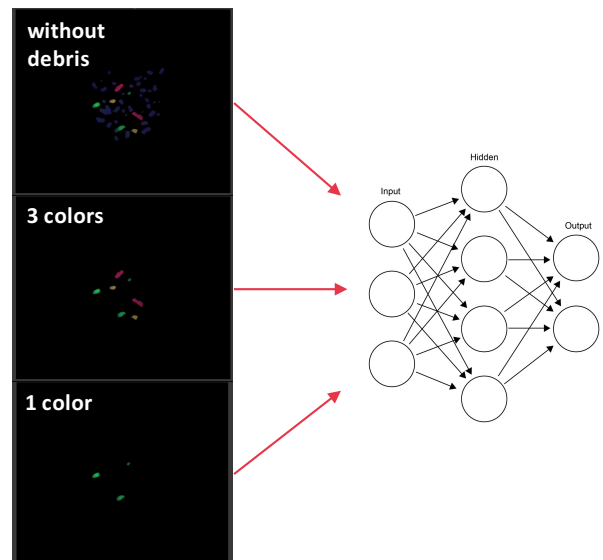
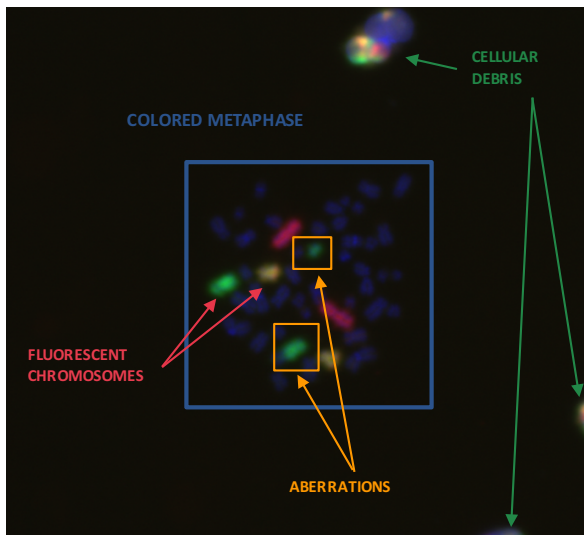
Aberrations visible thanks to the co-locations of colors so spatial coordinates are essential



Encoding of expertise

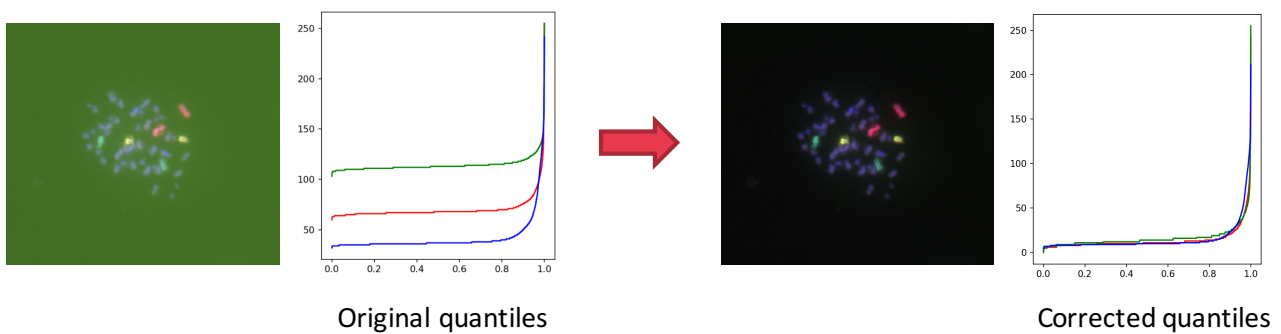


Transmitting information to AI models



Effort to homogenize FISH images

- **Homogenization** of data induces less confusion of vision models
- This corresponds to the **counting methodology** of experts in operational situations
- In particular, the correction of **excess fluorescence** (recalibration of the DAPI quantiles)



Contexte

Données

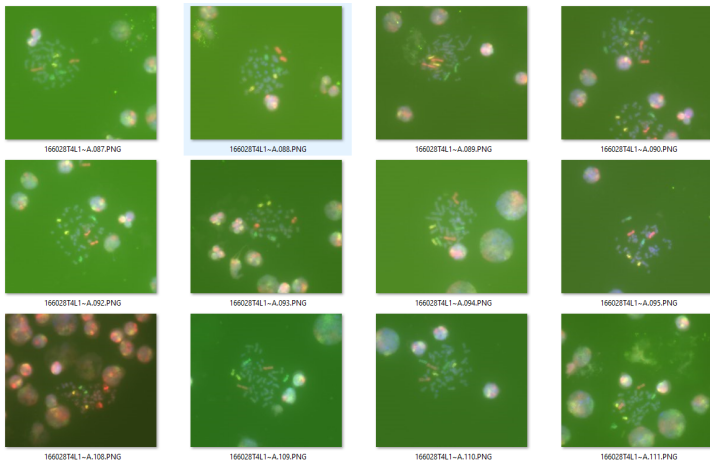
Métaphase

Chromosome

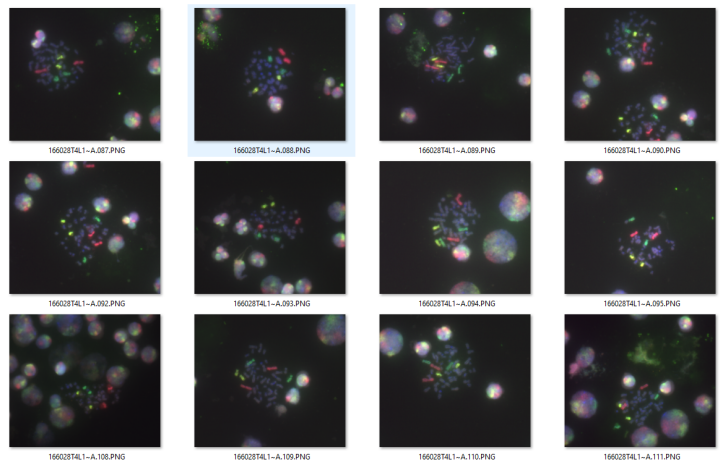
Génératif

Texture Correction Overview

BEFORE



AFTER



Contexte

Données

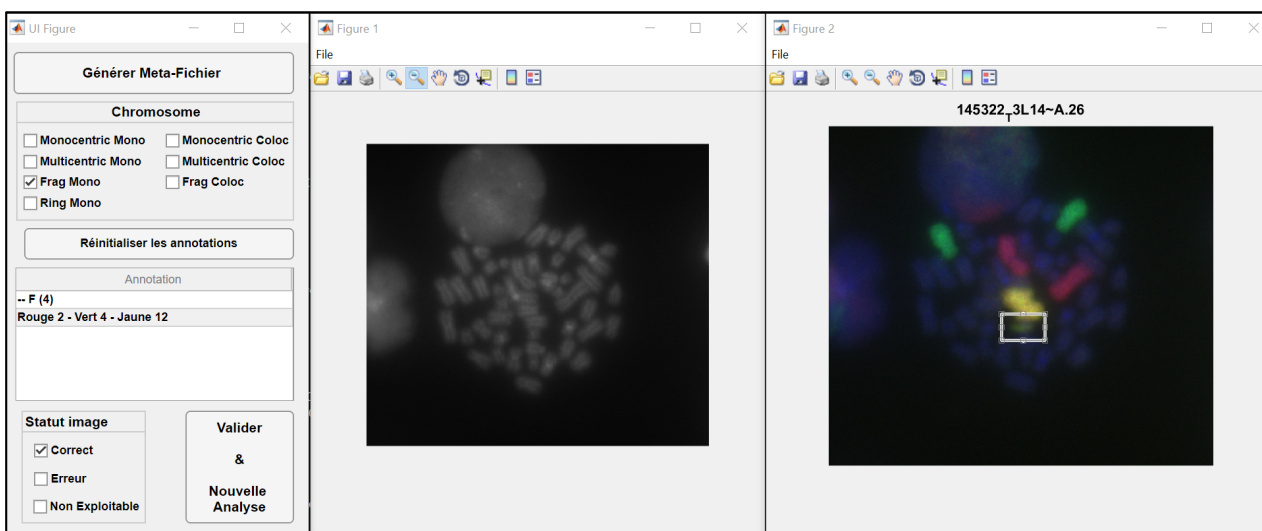
Métaphase

Chromosome

Génératif

Creation of the final database: FISH Audit

Adding spatial information to our data using **bounding boxes**



Contexte

Données

Métaphase

Chromosome

Génératif

Overview of the databases

	Images	Normal chromosomes	Translocated chromosomes	Aberrant chromosomes
DATASET 1 semi-supervised	357	2308	275 (11%)	N / A
DATASET 2 manual	452	2561	154 (5%)	328 (11%)
DATASET 3 manual FISH Audit	3476	18103	3123 (17%)	5962 (33%)

DATASET 1:

- annotations obtained in a **semi-supervised** manner
- only the **translocations** have been spatially annotated

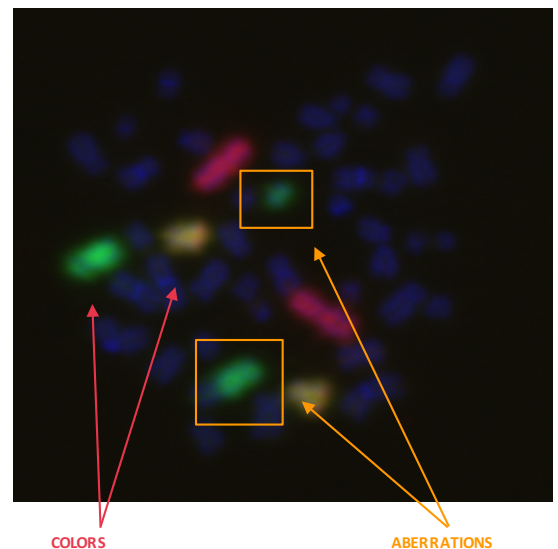
DATASET 2 & 3 (FISH Audit):

- **manual** annotations
- all the **aberrations** are spatially annotated

▮ **Imbalanced** classes: translocations are **underrepresented**

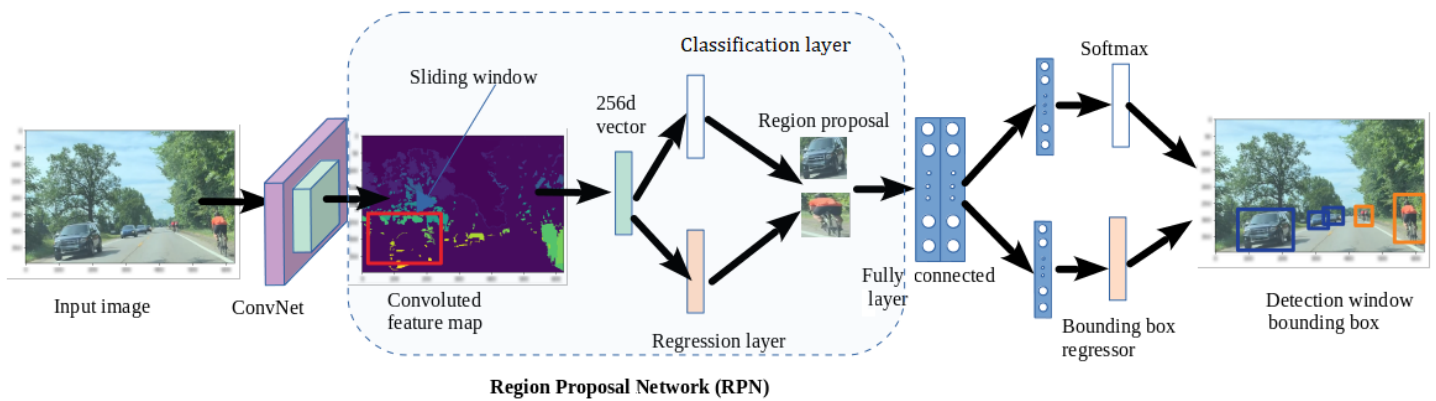
First detections at the metaphase scale

- The two detection tasks are to be solved **simultaneously**:
 - **Localization** of chromosomes of interest
 - **Classification** of localized chromosomes
- Nomenclature of the chromosome classes:
 - Class “**color**”: normal fluorescent chromosomes
 - Class “**aberration**”: aberrant chromosomes
- Supervised** object detection method: images and spatial annotations
 - 80% for **training and validation**
 - 20% for **testing**
- Evaluation of **model performances** on test data



State of the art in object detection: Faster-RCNN

Faster-RCNN uses **region proposal network (RPN)** to speed up the object detection process



Results obtained with Faster-RCNN

$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Recall} = \frac{tp}{tp + fn}$$

$$F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Class	Metric	DATASET 1 semi-supervised	DATASET 2 manual	DATASET 3 manual FISH Audit
Color	Precision	0.76	0.95	0.86
	Recall	0.77	0.96	0.94
	F1	0.76	0.95	0.90
Aberration	Precision	0.66	0.50	0.76
	Recall	0.23	0.33	0.67
	F1	0.35	0.40	0.71

- Similar scores for the Faster-RCNN on **GIEMSA Audit** (A. Deschemps PhD 2020-2023)
- Voting system with **heatmaps and model aggregation** is possible
- Strengthen decision making through the color class (6 colors in a normal metaphase)

Contexte

Données

Métaphase

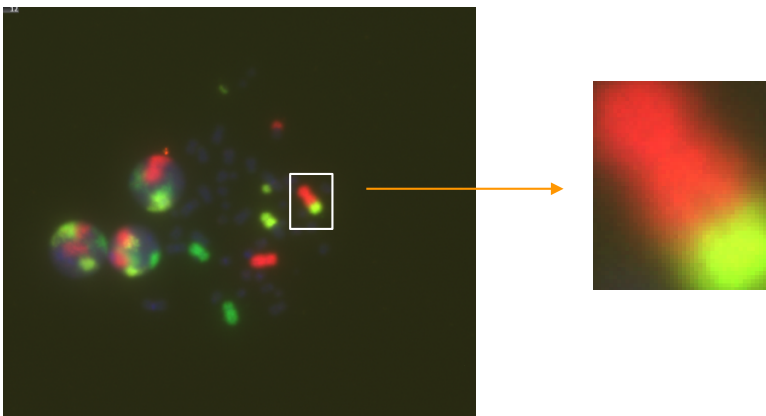
Chromosome

Génératif

Task separation to circumvent aberration rarity

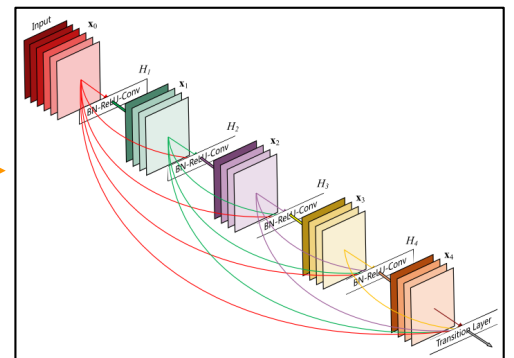
Step 1

Localization of chromosomes of interest



Step 2

Classification by convolutional networks such as ResNet



Contexte

Données

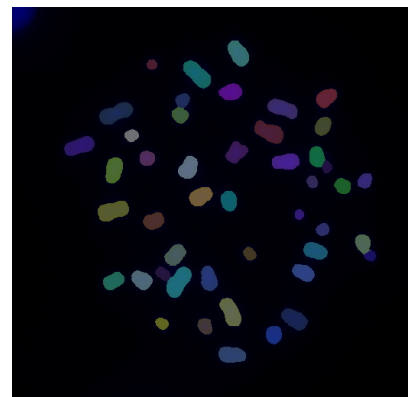
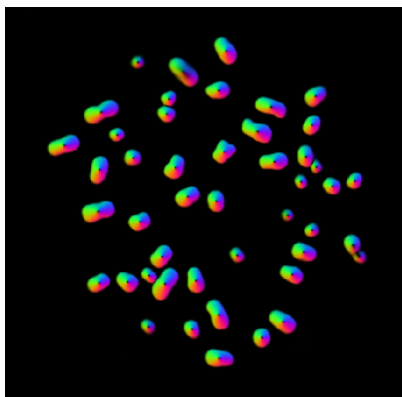
Métaphase

Chromosome

Génératif

Good segmentation allows work at the chromosome scale

- Cell nuclei segmentation model: **Cellpose**
 - Based on the **probability of presence** of a cell
 - Instance segmentation**: each chromosome is distinguished from the others
- Principle of **fine tuning**: adaptation of a model to new data



Contexte

Données

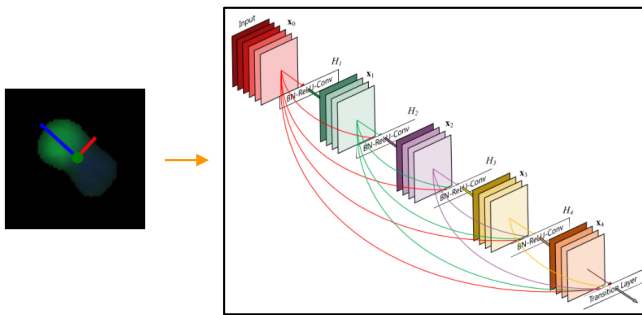
Métaphase

Chromosome

Génératif

Chromosome scale classification performance

ResNet: **residual convolutional** neural network



Class	Metric	ResNet (Chromosome)	Faster-RCNN (Metaphase)
Color	Precision	0.97	0.86
	Recall	0.97	0.94
	F1	0.97	0.90
Aberration	Precision	0.76	0.76
	Recall	0.70	0.67
	F1	0.73	0.71

Comparable performance between the metaphase scale and that of the chromosome which **validates data preprocessing** (segmentation+ homogenization)

Contexte

Données

Métaphase

Chromosome

Génératif

Exploring generative diffusion models

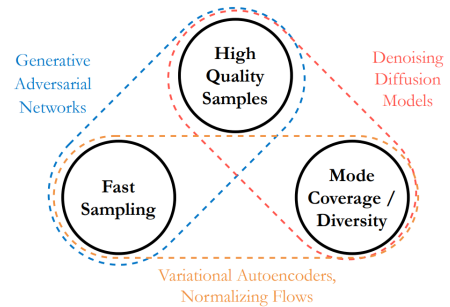
Reminder of the major challenges:

- **Imbalanced** classes
- **Little data** compared to the needs of the models

The generation of **synthetic translocations** is of major interest

Two possible cases of generation:

- Noise → Image
- **Image → Image**: possibility of DAPI input to generate fluorescence



Denoising Diffusion Probabilistic Models

Jonathan Ho
UC Berkeley
jonathanho@berkeley.edu

Ajay Jain
UC Berkeley
ajayj@berkeley.edu

Pieter Abbeel
UC Berkeley
pabbeel@cs.berkeley.edu

Brownian Bridge Diffusion Model (BBDM)

Model Image → Image:

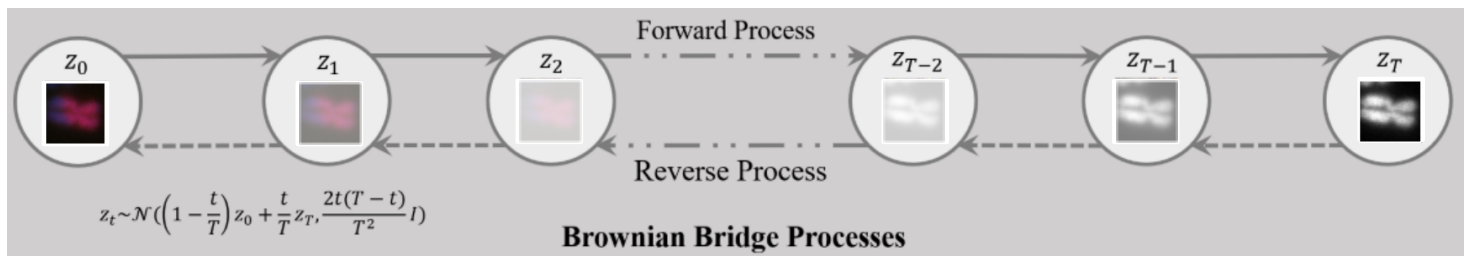
- Allows you to use the large amount of **unannotated DAPI**
- Reduces the generation to the **fluorescence**

Forward diffusion:

- Iterative corruption (Markov chain) of the original image Z_0 with gaussian noise

Backward/Reverse diffusion:

- Iterative de-corruption with neural network model (U-Net architecture)
- For a noisy image at the step Z_t , find the noise added in step $t - 1$



Contexte

Données

Métaphase

Chromosome

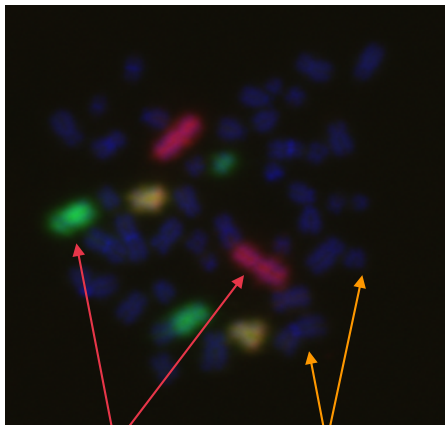
Génératif

Diffusion Model Test Pipeline

Generate chromosomes

Add to FISH Audit data

Compare ResNet scores



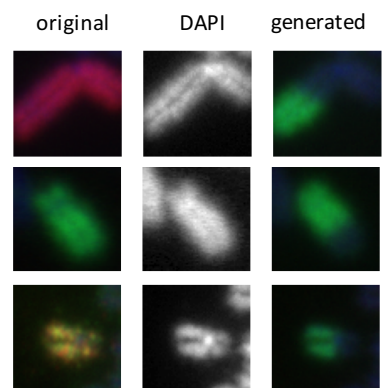
FLUORESCENT

NON-FLUORESCENT

Two generative databases:

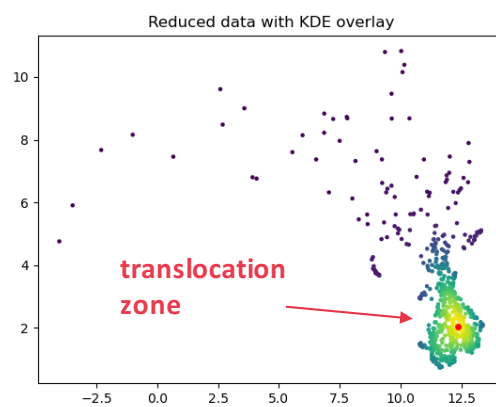
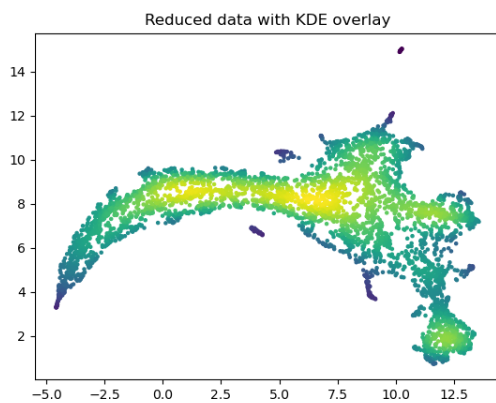
Data generated	Number of images
BASE 1: DAPI of fluorescent chromosomes	1154
BASE 2: DAPI of non-fluorescent chromosomes	10179

Examples of generations



Using ResNet intermediate layers for filtering

- An image is associated to a feature vector, **penultimate ResNet layer**
- **UMAP** dimension reduction (an image is represented by a point)



Translocations are concentrated

Contexte

Données

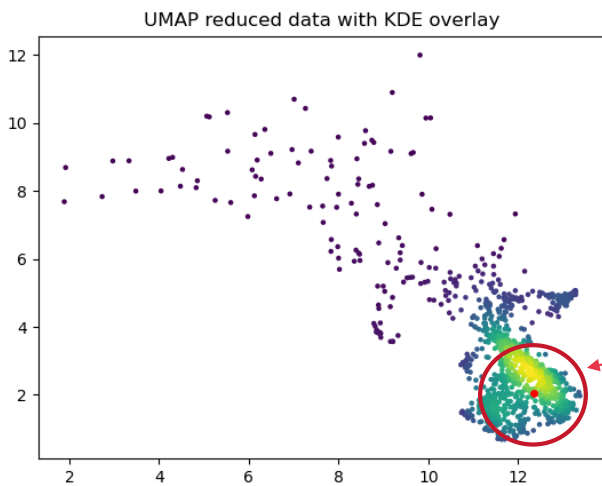
Métaphase

Chromosome

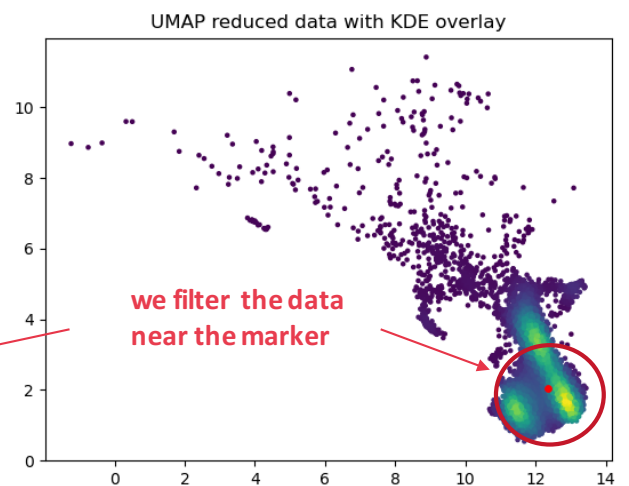
Génératif

UMAP Projection of Generative Databases

BASE 1: from DAPI of **fluorescent** chromosomes



BASE 2: from DAPI of **non-fluorescent** chromosomes



Data	BASE 1	BASE 2
Number of images filtered by UMAP ResNet	820	7518

Contexte

Données

Métaphase

Chromosome

Génératif

F1 scores of the aberration class using the generated data

BASE 1: from DAPI of **fluorescent** chromosomes

Case study	Score	Cardinal
Real data	73%	N = 1021
Generated Data	70%	N = 820
Real + Generated Data	80%	N = 1841

BASE 2: from DAPI of **non-fluorescent** chromosomes

Case study	Score	Cardinal
Real data	73%	N = 1021
Generated Data	68%	N = 7518
Real + Generated Data	76%	N = 8539

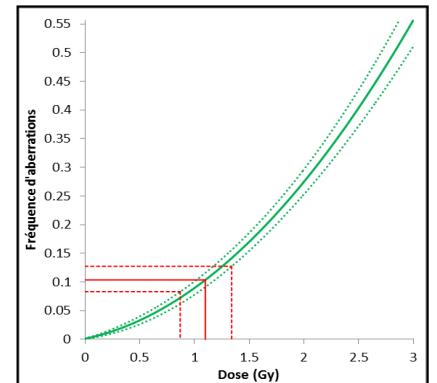
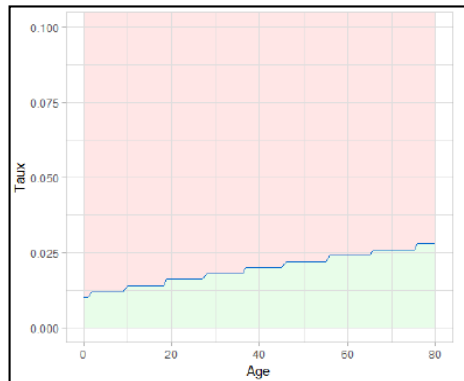
■ We must generate translocations from **morphologically close DAPI**

■ Allows us to reuse DAPI from fluorescent chromosomes of normal metaphases

Dosimetric Validation

- Construction of an **automatic counting tool** chromosomal translocations from standardized FISH images.
- Need to demonstrate implementation **operational work**: characterization **counting statistics** of the AI model and validation by **dose estimation**.

- Calibration curve modeling with base rate dependent on **demographic variables**



Contexte

Données

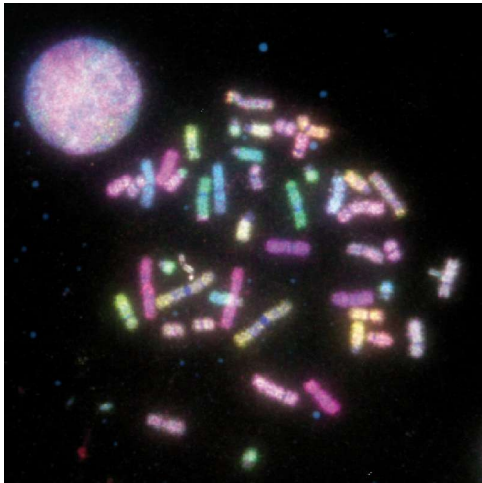
Métaphase

Chromosome

Génératif

Perspectives

| mFISH imagery greatly improves the quality of the data and thus the quality of the detection





**THANK YOU FOR YOUR
ATTENTION!**