

Projet Statistiques: Initiation à l'analyse des images en cosmologie et astro-physique

Hasna LAMMAIHRI, Omar KAMI, Abdellah BELLAHMAMA, Ahmed EL ALAOUI, Oussama CHAHBI

14 mai 2024

Encadré par M. Saad BENJELLOUN

Table des matières

1 Remerciement	2
2 Abstract	2
3 Introduction	2
3.1 Présentation du Télescope James Webb (JWST)	2
3.2 Importance du traitement d'images en astronomie	3
4 Dénombrément des étoiles dans les images du JWST	4
4.1 Contexte	4
4.1.1 Importance du dénombrement des étoiles	4
4.1.2 Défis de l'identification des étoiles	5
4.2 Méthodologie	5
4.2.1 Collection des données et des images	5
4.2.2 Techniques de prétraitement des images	6
4.2.3 Algorithmes d'identification des étoiles	8
4.2.4 Validation et analyse statistique	11
4.3 Résultats	12
4.3.1 Baseline Model et calcul de précision	15
4.4 Discussion des résultats	15
5 Estimation du mouvement propre des étoiles	16
5.1 Contexte	16
5.1.1 Paramètres astrométriques	16
5.1.2 Importance en astrophysique	17
5.2 Méthodologie	18
5.2.1 Data processing	18
5.2.2 Algorithmes de calcul	19
5.3 Visualisation des Résultats	22
5.3.1 Évolution du Mouvement Propre	22
5.3.2 Évolution de la Position Angulaire (PA) et de la Séparation (SEP) de l'Étoile	23
5.3.3 Conclusion :	24
6 Ouverture sur le Décryptage des Paramètres Morphologiques des Galaxies avec GaMPEN	25
6.1 Introduction	25
6.2 Implémentation Bayésienne de GaMPEN et Estimation des Incertitudes	26
6.3 Test de la Solution	29
6.3.1 Entrées et Sorties	29
7 Bibliographie	30
8 Annexe	31

1 Remerciement

Nous sommes profondément reconnaissants envers M. Saad BENJELLOUN, notre encadrant et professeur, pour leur dévouement et leur précieuse assistance tout au long de ce projet. Leur expertise, leur soutien et leurs conseils éclairés ont été cruciaux pour nous orienter à travers les défis techniques et conceptuels rencontrés lors de notre projet. Leur accompagnement attentif et les retours constructifs ont été d'une valeur inestimable, contribuant ainsi grandement à la réussite de ce projet.

Enfin, nous tenons à témoigner notre profonde gratitude envers notre école, l'EMINES School of Industrial Management, pour nous avoir donné l'occasion d'explorer un domaine aussi captivant que la statistique et de nous immerger dans des modèles complexes.

2 Abstract

Ce projet vise à exploiter les images du Télescope Spatial James Webb (JWST) pour deux objectifs principaux : le dénombrement des étoiles à l'aide de techniques de prétraitement (filtres, seuillage) et d'algorithmes de détection d'objets et de bords (Canny, Rectagle detection), et l'estimation des postérieurs pour une prédiction robuste des incertitudes. En intégrant des techniques avancées de traitement d'image et d'analyse statistique, nous aspirons à affiner les méthodes de quantification stellaire. Le project inclut également une perspective future de décryptage des paramètres morphologiques des galaxies en utilisant réseau de neurone convolutif GaMPEN, ouvrant ainsi la voie à des études plus approfondies sur la structure et l'évolution galactique.

3 Introduction

Ce projet avait pour objectif principal de nous initier à l'analyse des images en cosmologie et astrophysique, en exploitant les bibliothèques Python dédiées à cette fin. Nous nous étions concentrés spécifiquement sur l'utilisation des données provenant du télescope spatial James Webb (JWST), cherchant à reproduire partiellement certaines analyses présentées dans la littérature scientifique.

Le télescope James Webb représente une avancée majeure dans l'observation cosmique, offrant des opportunités sans précédent pour sonder les mystères de l'univers. En nous appuyant sur les données capturées par cet instrument révolutionnaire, notre projet visait à développer une compréhension approfondie des méthodes d'analyse d'images en astronomie, en mettant l'accent sur l'utilisation de librairies Python spécialisées.

À travers cette démarche, nous avions cherché à acquérir une expertise pratique dans le traitement et l'analyse des données astronomiques, tout en explorant les défis uniques posés par les images capturées par le JWST. En reproduisant partiellement des analyses de la littérature, nous souhaitions consolider notre compréhension des techniques utilisées dans la recherche cosmologique et contribuer modestement à l'avancement des connaissances dans ce domaine fascinant.

Ce rapport documente notre parcours à travers cette exploration, en détaillant nos méthodes, nos résultats et nos réflexions tout au long du processus. En mettant l'accent sur l'apprentissage pratique et l'application des concepts théoriques, nous avons cherché à fournir une ressource utile pour ceux qui souhaitent se familiariser avec l'analyse des images en cosmologie et astrophysique, en utilisant les outils puissants offerts par le langage de programmation Python.

3.1 Présentation du Télescope James Webb (JWST)

Lancé le 25 décembre 2021, JWST est le fruit d'une collaboration internationale impliquant la NASA, l'Agence spatiale européenne (ESA) et l'Agence spatiale canadienne (ASC). Doté d'un miroir principal de 6,5 mètres de diamètre et d'une suite d'instruments scientifiques avancés (FGS-NIRISS, MIRI, NIRCam, NIRSpec), le JWST est conçu pour observer l'univers dans l'infrarouge. Cette capacité lui permet de percer à travers les nuages de poussière cosmique et de révéler des régions de l'espace inaccessibles par les télescopes optiques.

Les objectifs scientifiques du JWST incluent la recherche des premières galaxies formées après le Big Bang, l'étude de la formation et de l'évolution des galaxies, l'observation des systèmes stellaires et planétaires en formation, ainsi que l'analyse des atmosphères des exoplanètes, potentiellement habitables.

Il opère près du point de Lagrange L2, à environ 1,5 million de kilomètres de la Terre, où il bénéficie d'un environnement thermiquement stable et d'une vue claire de l'espace. Sa conception

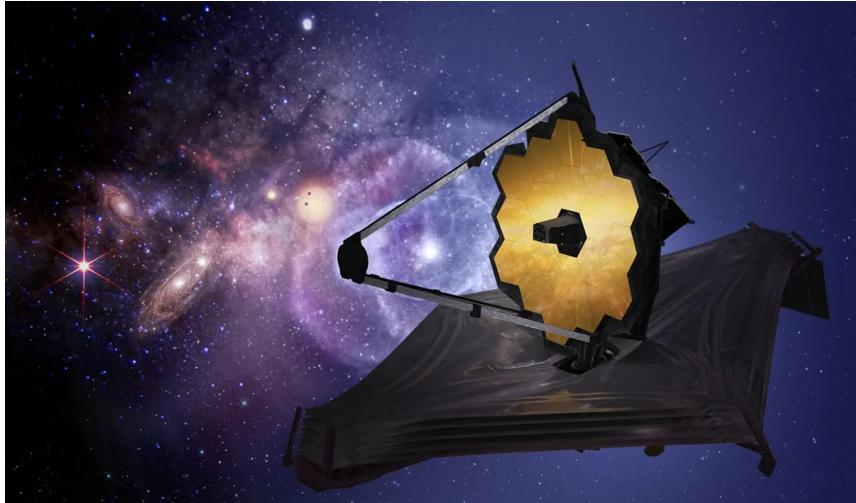


FIGURE 1 – james webb space telescope

unique avec un bouclier solaire de la taille d'un terrain de tennis lui permet de rester à l'ombre, maintenant ses instruments à des températures extrêmement basses nécessaires pour l'astronomie infrarouge.

Le JWST est équipé de quatre instruments scientifiques principaux, chacun conçu pour exploiter ses capacités d'observation dans l'infrarouge. Ces instruments permettent de réaliser une variété d'observations, de la spectroscopie à l'imagerie.

1. NIRCam (Near Infrared Camera)

La NIRCam est la principale caméra infrarouge du JWST et sert d'instrument crucial pour l'observation des étoiles, des galaxies, et d'autres corps célestes dans le proche infrarouge. Elle est conçue pour capter des images de haute résolution dans une plage de longueurs d'onde de 0,6 à 5 microns. La NIRCam est également essentielle pour les opérations de guidage du télescope et joue un rôle clé dans les études de formation étoile et de détection d'exoplanètes.

2. NIRSpec (Near Infrared Spectrograph)

Le NIRSpec permet la spectroscopie de plusieurs objets simultanément grâce à un dispositif à micro-obturateurs qui peut isoler jusqu'à 100 objets célestes dans son champ de vision. Cette capacité multiplexée est précieuse pour l'étude détaillée des propriétés chimiques et physiques des galaxies lointaines, des étoiles jeunes et des atmosphères d'exoplanètes.

3. MIRI (Mid-Infrared Instrument)

MIRI combine à la fois des capacités de caméra et de spectrographe dans le moyen infrarouge (de 5 à 28 microns). Elle est particulièrement sensible aux régions plus froides de l'univers, où des étoiles et des planètes se forment. MIRI est donc cruciale pour observer les phénomènes tels que les disques proto-planétaires, les nébuleuses, et les galaxies actives au-delà de la portée du NIRCam et du NIRSpec.

4. FGS/NIRISS (Fine Guidance Sensor and Near Infrared Imager and Slitless Spectrograph)

Le FGS est utilisé pour stabiliser le pointage du télescope avec une précision extrême, tandis que le NIRISS offre des capacités uniques pour l'imagerie et la spectroscopie sans fente. Ce dernier est particulièrement utile pour l'étude d'exoplanètes par transit et l'analyse spectrale de sources lumineuses individuelles dans des champs d'étoiles denses.

3.2 Importance du traitement d'images en astronomie

En astronomie, le traitement d'images est essentiel pour extraire et analyser les données recueillies par les télescopes comme le James Webb Space Telescope (JWST). Ce processus permet non seulement de clarifier les observations visuelles, mais aussi de découvrir des phénomènes cosmiques qui ne sont pas immédiatement apparents.

1. Amélioration de la Qualité des Images

Les images capturées par les instruments astronomiques sont souvent affectées par divers facteurs tels que le bruit de fond cosmique, les interférences instrumentales, et les distorsions optiques. Le traitement d'images permet de corriger ces défauts à travers des techniques telles que le filtrage, l'alignement des images, et la correction de l'étalement des étoiles (PSF). En résultat, les images deviennent plus nettes et les détails plus discernables, ce qui est crucial pour l'analyse précise de

régions d'intérêt spécifiques.

2. Détection et Classification des Objets Célestes

Les algorithmes de traitement d'images jouent un rôle clé dans la détection automatique des objets célestes tels que les étoiles, les galaxies, et les exoplanètes. Des techniques comme le seuillage, la segmentation, et la classification basée sur l'apprentissage machine permettent d'identifier et de cataloguer ces objets de manière systématique et répétable, fournissant une base de données riche pour des études plus approfondies.

3. Analyse Quantitative

Le traitement d'images permet d'extraire des quantités mesurables à partir des observations, telles que la luminosité, la taille, et le spectre des objets célestes. Ces mesures sont indispensables pour tester des théories astrophysiques, calculer les distances dans l'univers, et comprendre la structure et l'évolution des galaxies et des systèmes stellaires.

4. Maximisation des Ressources Observationnelles

Avec le coût élevé du temps télescope et les limitations imposées par les fenêtres d'observation, optimiser l'utilisation des données collectées est essentiel. Le traitement d'images permet d'exploiter pleinement chaque pixel d'information capturée par les télescopes, maximisant ainsi le retour scientifique des missions spatiales coûteuses.

5. Révélation de Phénomènes Subtils

Des techniques avancées comme le traitement par ondelettes, l'analyse de Fourier, et la cartographie des couleurs sont utilisées pour mettre en évidence des caractéristiques subtiles dans les données astronomiques, telles que les jets de matière éjectés par les trous noirs, les anneaux autour des planètes, et les structures filamenteuses dans les nuages moléculaires.

En conclusion, le traitement d'images est un pilier central de l'astronomie moderne. Il transforme les données brutes en un trésor de découvertes potentielles et joue un rôle crucial dans la façon dont nous percevons et comprenons l'univers. Grâce à ces techniques, des projets tels que ceux basés sur les images du JWST peuvent conduire à des percées significatives dans notre quête pour déchiffrer les mystères cosmiques.

4 Dénombrement des étoiles dans les images du JWST

4.1 Contexte

4.1.1 Importance du dénombrement des étoiles

Le dénombrement des étoiles offre des aspects cruciaux pour plusieurs domaines de recherche en astronomie et astrophysique.

1. Structure et Dynamique Galactique

Le dénombrement des étoiles aide les astronomes à comprendre la structure et la dynamique des galaxies, y compris la Voie Lactée. Par exemple, en analysant la distribution spatiale des étoiles, les chercheurs peuvent déduire la forme de la galaxie, identifier des bras spiraux, des halos stellaires, et des bulbes centraux. Ces informations sont essentielles pour modéliser l'évolution des galaxies et étudier les interactions galactiques.

2. Formation et Évolution des Étoiles

Les données obtenues par le dénombrement des étoiles permettent également d'étudier les régions de formation stellaire, en particulier dans les nuages moléculaires où de nouvelles étoiles sont en cours de formation. En comptant les étoiles dans différentes étapes de leur développement, les astronomes peuvent établir des modèles de formation stellaire et suivre l'évolution stellaire au fil du temps.

3. Paramètres Cosmologiques

Le dénombrement des étoiles contribue à l'évaluation de paramètres cosmologiques importants, tels que la densité stellaire de l'univers et la fonction de luminosité des galaxies. Ces mesures sont cruciales pour comprendre la distribution de la matière dans l'univers et pour tester des modèles cosmologiques qui expliquent l'expansion et l'évolution de l'univers.

4. Détection d'Exoplanètes

En comptant les étoiles et en analysant leur luminosité et leurs variations temporelles, les astronomes peuvent détecter des exoplanètes par la méthode des transits. Quand une planète passe devant son étoile hôte, elle cause une diminution légère mais mesurable de la luminosité de l'étoile, ce qui permet d'inférer l'existence et les propriétés de la planète.

5. Calibration des Distances Cosmiques

Le dénombrement des étoiles est également utilisé pour calibrer des échelles de distance dans l'univers. Par exemple, en identifiant des étoiles standard comme les céphéides, dont la luminosité

intrinsèque est bien connue, les astronomes peuvent mesurer les distances à d'autres galaxies, un élément fondamental pour cartographier l'expansion de l'univers.

En somme, le dénombrement des étoiles est une technique indispensable en astronomie. Elle permet non seulement de cataloguer les étoiles et d'étudier leur distribution, mais aussi de répondre à des questions plus larges concernant la structure de l'univers, la formation des étoiles et des galaxies, et l'évolution cosmique. En exploitant les images du JWST, cette méthode s'enrichit considérablement, promettant de nouvelles découvertes dans le domaine de l'astronomie.

4.1.2 Défis de l'identification des étoiles

L'identification précise des étoiles dans les images capturées par des télescopes comme le JWST est cruciale pour de nombreuses études astronomiques. Cependant, ce processus est confronté à divers défis qui peuvent affecter la qualité et la fiabilité des données obtenues.

1. Superposition et Encombrement Stellaires

Dans les régions denses de l'univers, comme les centres galactiques ou les amas d'étoiles, la superposition des étoiles est fréquente. Cette superposition rend difficile la distinction entre les sources lumineuses individuelles. L'encombrement stellaire peut conduire à des erreurs dans le dénombrement des étoiles et affecter la précision des mesures photométriques.

2. Variabilité Stellaires

Les étoiles ne sont pas des sources de lumière constantes ; beaucoup varient en luminosité en raison de phénomènes internes ou de facteurs externes comme les éclipses par des exoplanètes. Identifier et cataloguer les étoiles variables nécessite des observations répétées sur de longues périodes, ce qui complique le processus d'identification et exige des analyses temporelles complexes.

3. Limitations des DéTECTEURS

Les détecteurs utilisés dans l'astronomie ont leurs propres limites, y compris la sensibilité à certaines longueurs d'onde, la gamme dynamique, et le bruit de fond. Ces limitations peuvent empêcher la détection de sources faibles ou extrêmement éloignées et introduire des incertitudes dans l'identification des étoiles.

4. Difficultés de Calibration

La calibration des instruments est essentielle pour garantir l'exactitude des mesures photométriques et astrométriques. Des erreurs de calibration peuvent survenir en raison de changements dans les conditions instrumentales ou de défaillances dans les systèmes de référence. Ces erreurs peuvent fausser les estimations de la luminosité et de la position des étoiles.

5. Interférence des Objets de Fond

Dans certaines images, les galaxies, les quasars et d'autres objets non stellaires peuvent être confondus avec des étoiles, surtout si leur résolution n'est pas suffisante pour les résoudre clairement. Discriminer entre ces différents types d'objets nécessite des techniques de traitement d'image avancées et souvent des observations complémentaires.

6. Problématiques de Traitement d'Image

Les défis du traitement d'images, comme le débruitage, l'amélioration du contraste, et l'extraction de caractéristiques, sont essentiels pour une identification correcte des étoiles. Ces techniques doivent être continuellement affinées pour s'adapter aux caractéristiques uniques des données et aux nouvelles découvertes scientifiques.

4.2 Méthodologie

4.2.1 Collection des données et des images

Pour la collecte de données et d'images du Télescope Spatial James Webb (JWST), nous avons choisi d'utiliser le portail MAST (Mikulski Archive for Space Telescopes) pour accéder à ces données suivant les étapes suivantes :

1. Accéder au site-web du [portail MAST](#).
2. Entrer le nom de la cible (Target Name) : Exemple : TRAPPIST-1 , ainsi que les valeurs des différents filtres notamment les instruments utilisés, le type d'objet, la plage de dates..
3. Télécharger les données sous format FITS

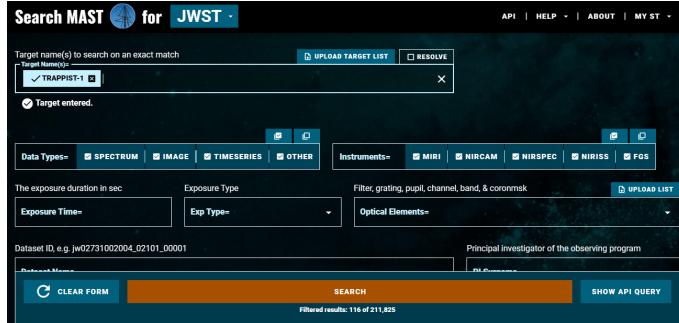


FIGURE 2 – Portail de recherche MAST 1

This screenshot shows the results page from the MAST search. The top bar includes "EDIT SEARCH" and "Target= trappist-1". It also shows "Columns= Dataset, Product Levels, Target Name, RA (J2000), Dec (J2000), Instrument, Exp Type, Optical Element...". The main area displays a table of data with columns: Level, Target Name, RA (J2000), Dec (J2000), Instrument, Exp Type, Optical Element, and Start Time. The table lists multiple entries for TRAPPIST-1, each with different coordinates and instrument details. At the bottom, it says "1 dataset selected".

FIGURE 3 – Portail de recherche MAST 2

4.2.2 Techniques de prétraitement des images

Pour mieux comprendre les propriétés inhérentes des images et les procédures techniques utilisées pour les manipuler et les traiter, nous pouvons considérer une image, composée de pixels individuels, comme une fonction f . Chaque pixel a sa propre valeur. Pour une image en niveaux de gris, chaque pixel aurait une intensité entre 0 et 255, où 0 est noir et 255 est blanc. $f(x,y)$ donnerait alors l'intensité de l'image à la position du pixel (x,y) , sachant qu'elle est définie sur un rectangle, avec une plage finie :

$$f : [a,b] \times [c,d] \rightarrow [0, 255]$$

Une image en couleur est simplement une extension de cela. $f(x,y)$ est maintenant un vecteur de trois valeurs au lieu d'une. En utilisant une image RGB comme exemple, les couleurs sont construites à partir d'une combinaison de Rouge, Vert et Bleu (RGB). Par conséquent, chaque pixel de l'image a trois canaux et est représenté comme un vecteur 1×3 . Puisque les trois couleurs ont des valeurs entières de 0 à 255, il y a un total de $256 \times 256 \times 3 = 16,777,216$ combinaisons ou choix de couleurs.

$$f(x,y) = \begin{bmatrix} r(x,y) \\ g(x,y) \\ b(x,y) \end{bmatrix}$$

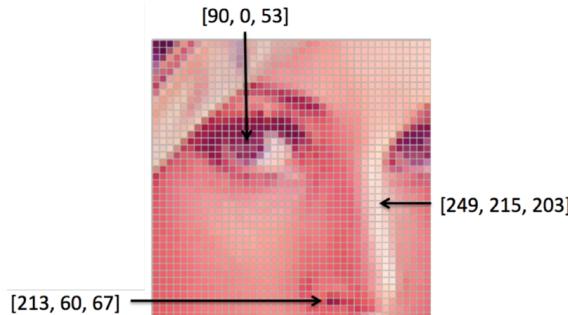


FIGURE 4 – De l'image à la matrice

Le filtrage est l'un des opérations de traitement d'images les plus utilisées. Un filtre est dit linéaire s'il :

- est invariant à la position : le résultat du filtrage ne dépend que du voisinage d'un pixel et non de la position absolue du pixel dans l'image
- respecte les principes de mise à l'échelle et de superposition (le résultat du filtrage de la somme de deux images est le même que celui de la somme des deux images filtrées) :

$$f(ax + by) = af(x) + bf(y)$$

Or, on sait qu'une multiplication dans le domaine des fréquences est équivalente à une convolution dans le domaine spatial, c'est-à-dire :

$$f(t) * s(t) \Leftrightarrow F(\omega)S(\omega)$$

Cet opérateur de convolution prend la forme d'un masque ou noyau ("kernel") de convolution. L'opération de filtrage consiste à convoluer ce masque avec l'image.

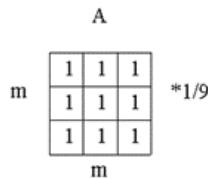


FIGURE 5 – Exemple : Filtre moyenneur 3 x 3

La convolution consiste à balayer l'intégralité de l'image avec le masque :

$$E_{\text{filtre}}(i, j) = A \cdot E(i, j) = \frac{1}{m^2} \sum_{k=-\frac{m}{2}}^{\frac{m}{2}} \sum_{h=-\frac{m}{2}}^{\frac{m}{2}} A(h, k)E(i - h, j - k)$$

Remarque : La division par le nombre de pixels du kernel (m^2) n'appartient pas au produit de convolution proprement dit, mais elle est nécessaire pour maintenir la dynamique de l'image ainsi que sa linéarité.

Filtre Gaussien Le filtre gaussien est un filtre linéaire dont le noyau de convolution est de forme gaussienne définie par :

$$G_\sigma(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

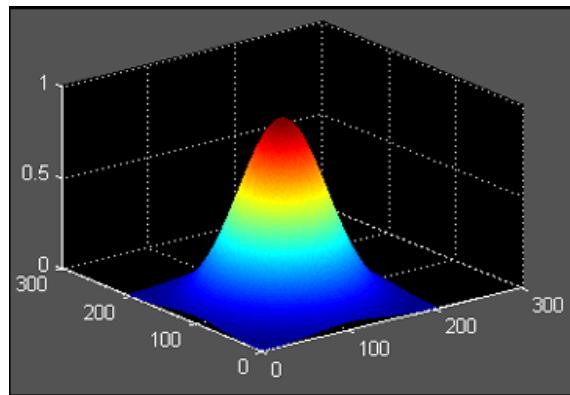


FIGURE 6 – Enter Caption

Filtre median Est un filtre non linéaire qui utilise un noyau sur lequel on effectue les opérations suivantes :

1. trier les valeurs d'illuminance des pixels couverts par le masque
2. extraire la médiane des données triées
3. remplacer la valeur du pixel central par la médiane

Seuillage d’Otsu Il s’agit d’une méthode couramment utilisée en Computer Vision et en traitement d’images pour effectuer un seuillage automatique des images. L’algorithme donne un seul seuil d’intensité qui sépare les pixels en deux classes : premier plan et arrière-plan. Le seuil est déterminé en minimisant la variance des intensités intra-classe (d’une manière équivalente, en maximisant la variance inter-classes) La méthode d’Otsu est une version unidimensionnelle discrète de l’analyse discriminante de Fisher, c’est liée à la méthode d’optimisation de Jenks, et est équivalente à un K-means optimal global réalisé sur l’histogramme d’intensité.

L’algorithme cherche de manière exhaustive le seuil qui minimise la variance intra-classe, définie comme une somme pondérée des variances des deux classes : $\sigma_w^2(t) = \omega_0(t)\sigma_0^2(t) + \omega_1(t)\sigma_1^2(t)$ Où ω_i est la probabilité de la classe i séparée par le seuil t , et σ_i^2 est la variance de la classe i . La probabilité $\omega_i(t)$ est calculée à partir des L bins de l’histogramme : $\omega_0(t) = \sum_{i=0}^{t-1} p(i)$ et $\omega_1(t) = \sum_{i=t}^{L-1} p(i)$

Pour deux classes, minimiser la variance intra-classe revient à maximiser la variance inter-classes. $\sigma_b^2(t) = \sigma^2 - \sigma_w^2(t) = \omega_0(t)(\mu_0 - \mu_T)^2 + \omega_1(t)(\mu_1 - \mu_T)^2$

$$= \omega_0(t)\omega_1(t)[\mu_0(t) - \mu_1(t)]^2$$

Ceci est exprimé en termes de probabilités de classe ω et de moyennes de classe μ :

$$\mu_0(t) = \frac{\sum_{i=0}^{t-1} ip(i)}{\omega_0(t)}, \mu_1(t) = \frac{\sum_{i=t}^{L-1} ip(i)}{\omega_1(t)} \text{ et } \mu_T = \sum_{i=0}^{L-1} ip(i)$$

Les relations suivantes peuvent être facilement vérifiées : $\omega_0\mu_1 + \omega_1\mu_1 = \mu_T$
 $\omega_0 + \omega_1 = 1$

Algorithme :

- Calculer l’histogramme et les probabilités de chaque niveau d’intensité
- Initialiser $\omega_i(0)$ et $\mu_i(0)$
- Parcourir tous les seuils possibles $t = 1, \dots$, intensité maximale
 - Mettre à jour ω_i et μ_i
 - Calculer $\sigma_b^2(t)$
- Le seuil souhaité correspond au maximum de $\sigma_b^2(t)$

4.2.3 Algorithmes d’identification des étoiles

Canny edge detection : C’est un opérateur de détection de contours qui utilise un algorithme multi-étapes pour détecter une large gamme de contours dans les images. Il s’agit d’une technique permettant d’extraire des informations structurelles utiles à partir d’objets de vision différents et de réduire de manière significative la quantité de données à traiter. Elle a été largement appliquée dans divers systèmes de vision par ordinateur. Les critères généraux pour la détection de contours sont les suivants :

1. Détection du contour avec un faible taux d’erreur, ce qui signifie que la détection doit capturer aussi précisément que possible autant de contours que possible dans l’image.
2. Le point de contour détecté par l’opérateur doit être localisé de manière précise au centre du contour.
3. Un contour donné dans l’image ne doit être marqué qu’une seule fois, et dans la mesure du possible, le bruit de l’image ne doit pas créer de faux contours.

Il s’agit de l’une des méthodes les plus strictement définies qui offre une détection bonne et fiable.

Il s’agit de l’une des méthodes les plus strictement définies qui offre une détection bonne et fiable.

Processus :

1. Appliquer un "filtre gaussien" pour lisser l’image afin de supprimer le bruit :

Étant donné que tous les résultats de détection de bord sont facilement affectés par le bruit dans l’image, il est essentiel de filtrer le bruit pour éviter toute détection erronée causée par celui-ci. Pour lisser l’image, un noyau de filtre gaussien est convolué avec l’image. Cette étape permet de légèrement lisser l’image afin de réduire les effets du bruit évident sur le détecteur de bord. L’équation pour un noyau de filtre gaussien de taille $(2k+1) \times (2k+1)$ est donnée par :

$$H_{ij} = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(i-(k+1))^2 + (j-(k+1))^2}{2\sigma^2}\right); 1 \leq i, j \leq (2k+1)$$

Il est important de comprendre que le choix de la taille du noyau gaussien affectera les performances du détecteur ; plus la taille est grande, moins le détecteur sera sensible au bruit. De plus, l’erreur de localisation pour détecter le bord augmentera légèrement avec

la taille du noyau du filtre gaussien. Une taille de 5×5 est généralement recommandée pour la plupart des cas, mais cela peut également varier en fonction de situations spécifiques. Exemple de filtre gaussien 5×5 utilisé pour créer l'image adjacente, avec $\sigma = 1$ (l'astérisque

$$\text{indique une opération de convolution) : } B = \frac{1}{159} \begin{bmatrix} 2 & 4 & 5 & 4 & 2 \\ 4 & 9 & 12 & 9 & 4 \\ 5 & 12 & 15 & 12 & 5 \\ 4 & 9 & 12 & 9 & 4 \\ 2 & 4 & 5 & 4 & 2 \end{bmatrix} * A$$

2. Trouver les gradients d'intensité de l'image :

L'étape de calcul du gradient détecte l'intensité et la direction des bords en calculant le gradient de l'image à l'aide d'opérateurs de détection de bords. Les bords correspondent à un changement d'intensité des pixels. Pour les détecter, la manière la plus simple est d'appliquer des filtres qui mettent en évidence ce changement d'intensité dans les deux directions : horizontale (x) et verticale (y). Lorsque l'image est lissée, les dérivées I_x et I_y par rapport à x et y sont calculées. Cela peut être mis en œuvre en convoluant I avec les

$$\text{noyaux Sobel } K_x \text{ et } K_y, \text{ respectivement : } K_x = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}, K_y = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix}$$

Ensuite, la magnitude G et la pente θ du gradient sont calculées comme suit :

$$|G| = \sqrt{I_x^2 + I_y^2}$$

$$\theta(x, y) = \arctan\left(\frac{I_y}{I_x}\right)$$

3. Suppression des non-maximum :

Idéalement, l'image finale devrait avoir des bords fins. Ainsi, nous devons effectuer une suppression des non-maximum pour affiner les bords.

Le principe est simple : l'algorithme parcourt tous les points de la matrice d'intensité du gradient et trouve les pixels ayant la valeur maximum dans les directions des bords.

Exemple :

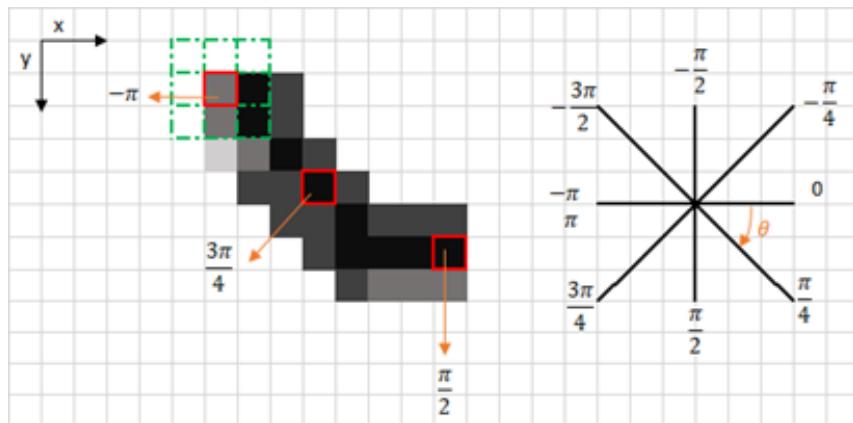


FIGURE 7 – Exemple suppression de non-maximum (1)

Le coin supérieur gauche de la cellule rouge présente sur l'image ci-dessus représente un pixel d'intensité de la matrice du gradient d'intensité en cours de traitement. La direction du bord correspondant est représentée par la flèche orange avec un angle de $-\pi$.

La direction du bord est représentée par la ligne en pointillés orange (horizontale de gauche à droite). L'objectif de l'algorithme est de vérifier si les pixels dans la même direction sont plus ou moins intenses que ceux en cours de traitement. Dans l'exemple ci-dessus, le pixel (i, j) est en cours de traitement, et les pixels dans la même direction sont mis en évidence en bleu $(i, j - 1)$ et $(i, j + 1)$. Si l'un de ces deux pixels est plus intense que celui en cours de traitement, alors seul le plus intense est conservé. Le pixel $(i, j - 1)$ semble être plus intense, car il est blanc (valeur de 255). Ainsi, la valeur d'intensité du pixel actuel (i, j) est remis à zéro. S'il n'y a pas de pixels dans la direction du bord ayant des valeurs plus intenses, alors la valeur du pixel actuel est conservée.

Dans ce cas, la direction est la ligne diagonale en pointillés orange. Par conséquent, le pixel le plus intense dans cette direction est le pixel $(i - 1, j + 1)$. Pour récapituler, chaque

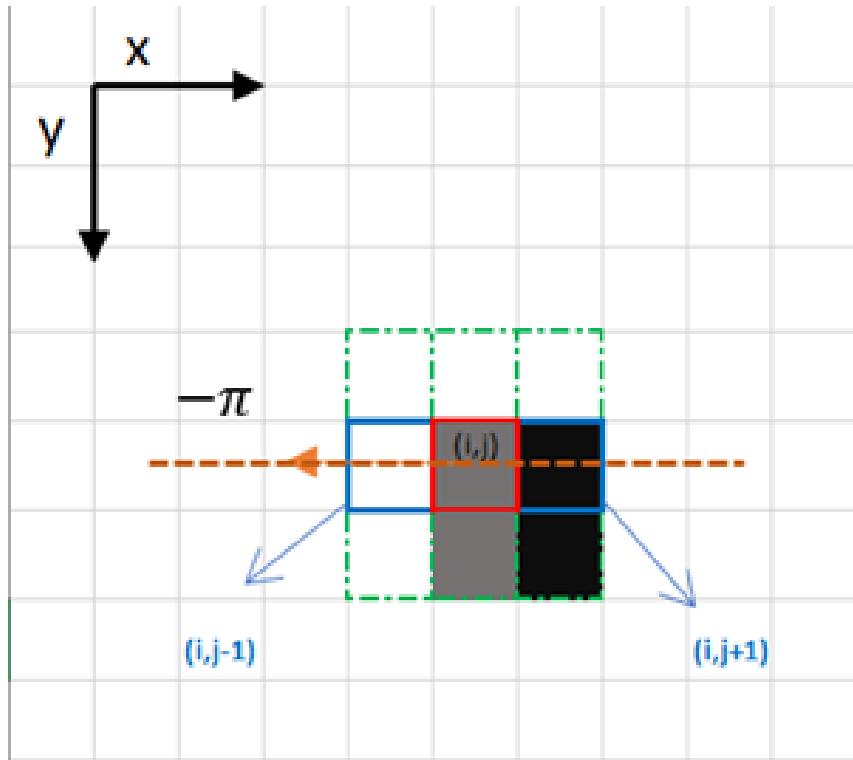


FIGURE 8 – Exemple suppression de non-maximum (2)

pixel possède deux critères principaux (la direction du bord et l'intensité du pixel, comprise entre 0 et 255). En fonction de ces critères, les étapes de la suppression des non-maximum sont les suivantes :

- Créer une matrice initialisée à 0 de la même taille que la matrice d'intensité du gradient d'origine.
 - Identifier la direction du bord en fonction de la valeur de l'angle dans la matrice des angles.
 - Vérifier si le pixel dans la même direction a une intensité plus élevée que le pixel actuellement traité.
 - Retourner l'image traitée avec l'algorithme de suppression des non-maximum.
4. Appliquer un double seuillage pour déterminer les contours potentiels L'étape du double seuillage vise à identifier trois types de pixels : forts, faibles et non pertinents :
 - Les pixels forts sont ceux dont l'intensité est si élevée que nous sommes sûrs qu'ils contribuent au bord final.
 - Les pixels faibles sont ceux dont la valeur d'intensité n'est pas suffisamment élevée pour être considérés comme forts, mais pas assez faible pour être considérés comme non pertinents pour la détection des bords.
 - Les autres pixels sont considérés comme non pertinents pour le bord.
- Les double seuils s'appliquent comme suit :
- Le seuil élevé est utilisé pour identifier les pixels forts (intensité supérieure au seuil élevé).
 - Le seuil bas est utilisé pour identifier les pixels non pertinents (intensité inférieure au seuil bas).
 - Tous les pixels ayant une intensité entre les deux seuils sont marqués comme faibles et le mécanisme d'hystérosis (étape suivante) nous aidera à identifier ceux qui pourraient être considérés comme forts et ceux qui sont considérés comme non pertinents.
5. Suivre les contours par "hystérosis". Finaliser la détection des contours en supprimant tous les autres contours qui sont faibles et ne sont pas connectés à des contours forts.
- Sur la base des résultats des seuils, l'hystérosis consiste à transformer les pixels faibles en pixels forts, si et seulement si au moins l'un des pixels autour de celui en cours de traitement est un pixel fort, comme décrit ci-dessous :

Rectangle detection

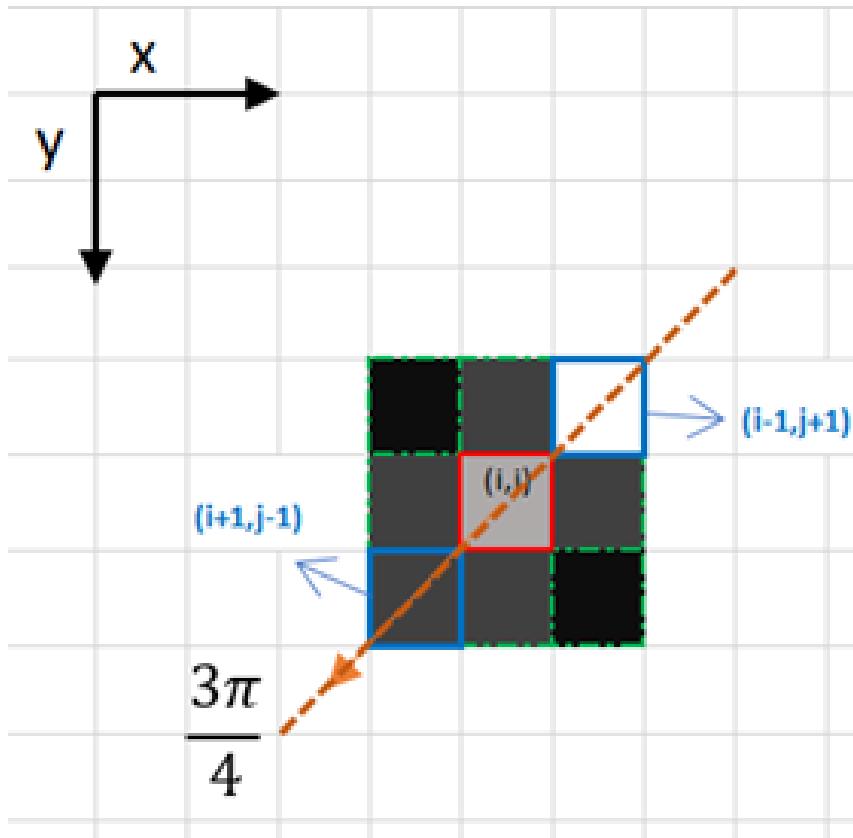


FIGURE 9 – Exemple suppression de non-maximum (3)

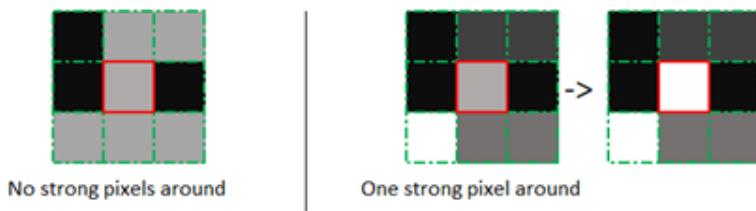


FIGURE 10 – Suivre les contours par hystérésis

4.2.4 Validation et analyse statistique

Anderson-Darling test : C'est un test statistique visant à déterminer si un échantillon de données donné provient d'une distribution de probabilité donnée. Dans sa forme de base, le test suppose qu'il n'y a pas de paramètres à estimer dans la distribution testée, auquel cas le test et son ensemble de valeurs critiques sont indépendants de la distribution. Cependant, le test est le plus souvent utilisé dans des contextes où une famille de distributions est testée, auquel cas les paramètres de cette famille doivent être estimés et cela doit être pris en compte dans l'ajustement soit de la statistique du test, soit de ses valeurs critiques. Lorsqu'il est appliqué pour déterminer si une distribution normale décrit de manière adéquate un ensemble de données, il est l'un des outils statistiques les plus puissants pour détecter la plupart des écarts par rapport à la normalité. Des tests d'Anderson-Darling pour k échantillons sont disponibles pour déterminer si plusieurs ensembles d'observations peuvent être modélisés comme provenant d'une seule population, sans que la fonction de distribution n'ait à être spécifiée. En plus de son utilisation comme test d'ajustement pour les distributions, il peut être utilisé dans l'estimation des paramètres comme base pour une forme de procédure d'estimation de la distance minimale. Le test d'Anderson-Darling est défini comme suit :

H_0 : Les données suivent une distribution spécifiée.

H_a : Les données ne suivent pas la distribution spécifiée.

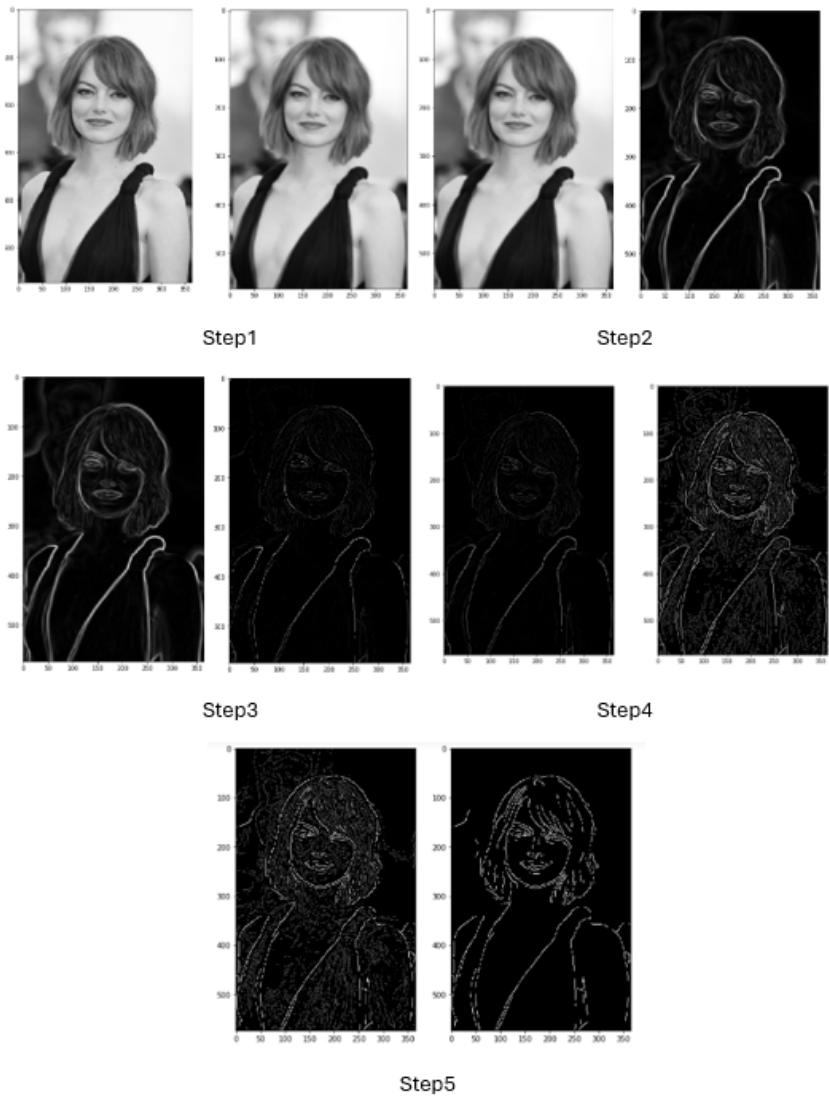


FIGURE 11 – Canny Edge Detection étape par étape

Statistique du test : La statistique du test d'Anderson-Darling est définie comme suit : $A^2 = -N - S$
Avec : $S = \sum_{i=1}^N \frac{(2i-1)}{N} [\ln F(Y_i) + \ln (1 - F(Y_{N+1-i}))]$

F est la fonction de distribution cumulative de la distribution spécifiée. Notez que les Y_i sont les données ordonnées. Les valeurs critiques pour le test Anderson-Darling dépendent de la distribution spécifique qui est testée. Des valeurs tabulées et des formules ont été publiées [Voir : Stephens 1974, 1976, 1977] pour quelques distributions spécifiques (normale, log-normale, exponentielle, de Weibull, logistique, de type valeur extrême 1 et autres). Le test est un test unilatéral et l'hypothèse selon laquelle la distribution est d'une forme spécifique est rejetée si la statistique de test, A^2 , est supérieure à la valeur critique. Notez que pour une distribution donnée, la statistique d'Anderson-Darling peut être multipliée par une constante (qui dépend généralement de la taille de l'échantillon, n). Ces constantes sont données dans les différents articles de Stephens. Dans la sortie d'échantillon ci-dessous, les valeurs de la statistique de test sont ajustées. De plus, veuillez noter que différentes constantes (et donc différentes valeurs critiques) ont été publiées. Vous devez simplement être conscient de la constante qui a été utilisée pour un ensemble donné de valeurs critiques (la constante nécessaire est généralement donnée avec les valeurs critiques).

4.3 Résultats

Dans cette partie, on applique les techniques précédentes pour compter les étoiles dans l'image célèbre "Pillars of Creation". Remarque : On considère que chaque objet lumineux dans l'image est une étoile, une étude plus rigoureuse nécessitera une classification de ces objets (étoile, galaxie, planète..) au préalable.

Dans un premier temps, on n'utilise que le seuillage d'Otsu :

Téléchargement des Images Nous avons choisi les célèbres "Pillars of Creation" dans la nébuleuse de l'Aigle (M16) comme sujet de notre étude. L'image a été téléchargée en format FITS depuis le portail MAST



FIGURE 12 – Image des "Pillars of Creation", JWST, 2022

Conversion de format Les fichiers FITS ont été convertis en images PNG à l'aide de la librairie AstroPy pour faciliter le traitement avec la librairie de vision artificielle OpenCV. Une portion carrée de l'image a été choisi pour diminuer le temps de calcul



Conversion en nuance de gris et Application du Filtre Médian Nous avons transformé les images en nuances de gris. Le filtre médian a été appliqué pour réduire le bruit de type "sel et poivre" tout en préservant les bords des structures stellaires, ce qui est crucial pour un dénombrement précis des étoiles.



Seuillage par la Méthode d’Otsu Nous avons utilisé le seuillage d’Otsu, segmentant ainsi les étoiles des fonds célestes. On a expérimenté avec différents valeurs de « threshold ».



Identification des Étoiles Les pixels excédant le seuil déterminé par Otsu ont été encadrés par un cercle rouge de rayon 1 pixel. Les régions rouges sont dus à plusieurs cercles proches confondus. On obtient un résultat de **1687719 étoiles**.

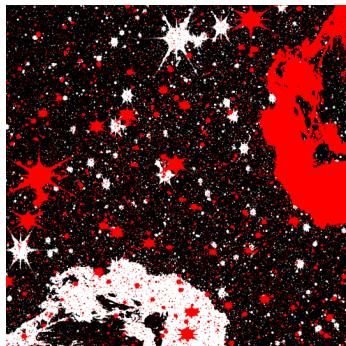


FIGURE 13 – Image après seuillage

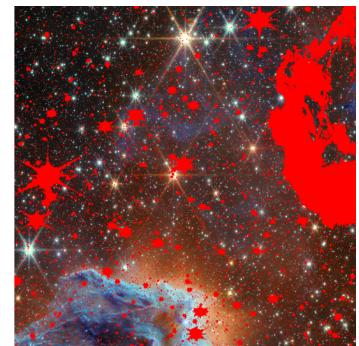


FIGURE 14 – Image originale à guise de compréhension

Pour améliorer la détection, on essaie, en plus du seuillage Otsu, les algorithmes de détection mentionnés précédemment. On reprend l’étude précédente :

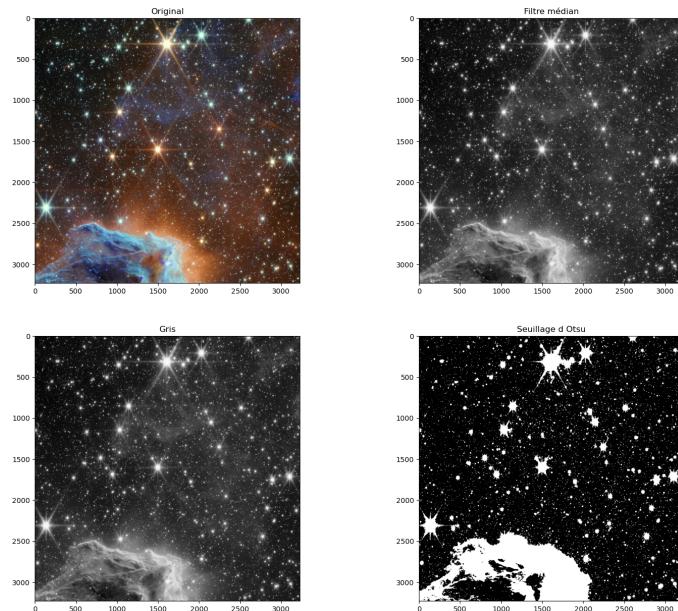
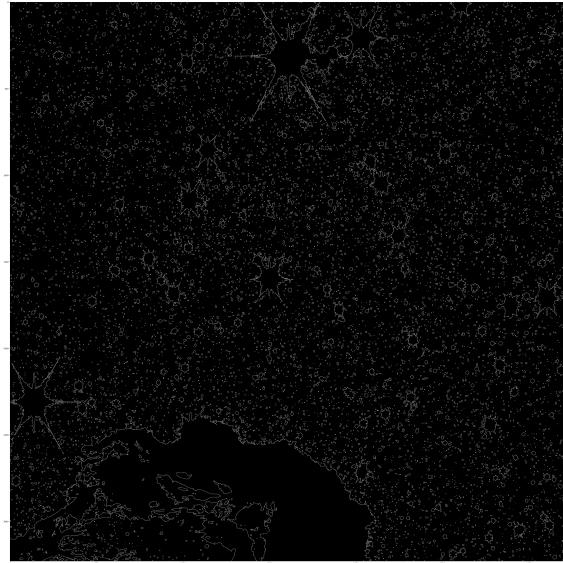


FIGURE 15 – Les différentes étapes de prétraitement

Rectangle Detection : Nous avons appliqué l'algorithme de détection de rectangles en utilisant la fonction "cv2.findContours" de la bibliothèque OpenCV, spécifiquement avec les paramètres "cv2.RETR-EXTERNAL" et "cv2.CHAIN-APPROX-SIMPLE". Cette méthode nous a permis d'identifier les contours extérieurs des étoiles de manière simplifiée, résultant en la détection de **1026 étoiles** pour cette image.

Canny Edge Detection : Ensuite, nous avons testé l'algorithme de détection de bords de Canny, une technique robuste pour identifier les changements d'intensité dans une image. En utilisant cette méthode, nous avons obtenu un résultat significativement plus élevé, avec **25281 étoiles** détectées. Cet écart s'explique par la sensibilité accrue de l'algorithme de Canny à capturer les détails fins dans les structures des étoiles.



4.3.1 Baseline Model et calcul de précision

Pour comparer les résultats et évaluer les performances de différentes méthodes, on a simulé des images astronomiques en générant des images de fond noir contenant des points blancs représentant les étoiles. Chaque image est de même dimension que l'image cible étudié ("Pillars of Creation) (3230 x 3230 pixels) et contient X étoiles qu'on varie à chaque itération. On varie X le nombre d'étoiles aléatoirement entre 4000 et 100000 et on applique à chaque fois chacune des méthodes précédentes. En calculant la moyenne on obtient les précisions dans le tableau 16.

$$X \in [4000, 100000]$$

$$\text{Précision} = \frac{1}{N} \sum_{i=1}^N \left(\frac{\text{Nombre d'étoiles détectées}_i}{\text{Nombre d'étoiles réelles}_i} \right) = \frac{1}{N} \sum_{i=1}^N \left(\frac{\text{Nombre d'étoiles détectées}_i}{X_i} \right)$$

où $N = 1000$ représente le nombre d'itérations et i l'indice de chaque itération.

4.4 Discussion des résultats

Méthode	Seuillage d'Otsu	Seuillage d'Otsu + Rectangle Detection	Seuillage d'Otsu + Canny edge detection
Résultat sur "Pillars of Creation"	1687719	1026	25281
Précision (Baseline)	~97%	~77%	~85%

FIGURE 16 – Comparaison des résultats des différentes méthodes de détection sur "Pillars of Creation"

Les résultats obtenus par les différentes méthodes de détection d'étoiles varient considérablement, à priori du au défis 4.1.2 de l'analyse d'images en astronomie, ce qui souligne l'importance

de choisir une technique adaptée au contexte spécifique de l'analyse d'image.

La méthode du seuillage d'Otsu a produit une surestimation significative sur l'image cible, malgré la variation des valeurs du "threshold", avec 1,687,719 étoiles détectées, principalement due à la difficulté de distinguer les étoiles de la poussière galactique et d'autres objets de fond. En revanche elle a produit une excellente détection sur l'image synthétique, ce qui est naturelle puisque cette méthode consiste à détecter toutes les pixels excédant le seuil.

En revanche, la détection de rectangles a sous-estimé le nombre d'étoiles, n'en identifiant que 1026, probablement en omettant les petites étoiles ou celles moins distinctes, en outre, elles est moins performante en terme de précision sur l'image synthétique.

La méthode de détection des bords de Canny semble offrir le meilleur équilibre, à la fois en terme de résultats sur l'image cible et en précision sur le baseline, avec 25,281 étoiles et une précision de 85% . Cette technique a montré une plus grande précision pour identifier les contours des étoiles sans confondre autant les étoiles avec le bruit de fond ou les autres éléments galactiques. Ce résultat suggère que la détection de Canny pourrait être plus adaptée pour les analyses où la distinction entre les étoiles et les objets de fond est cruciale.

Pour améliorer la détection des étoiles dans les images astronomiques, il est essentiel d'optimiser les techniques de prétraitement pour mieux distinguer les étoiles du bruit de fond. L'adoption de méthodes avancées comme l'apprentissage profond peut aider à classifier avec précision les étoiles. Combinaison de plusieurs méthodes de détection et l'utilisation de données multispectrales peuvent également enrichir l'analyse. Enfin, la mise en place d'un système de validation pour ajuster et affiner les algorithmes en fonction des résultats obtenus est cruciale pour garantir une détection efficace et précise.

5 Estimation du mouvement propre des étoiles

5.1 Contexte

5.1.1 Paramètres astrométriques

Dans cette section, nous aborderons la modélisation du mouvement prévu d'un objet céleste par rapport à une étoile cible donnée au fil du temps, en utilisant des données astronomiques telles que la parallaxe, les mouvements propres et les observations astrométriques. Avant d'entrer dans les détails de notre analyse, il est essentiel de comprendre la signification et l'importance de ces données. Voici les définitions succinctes de chaque paramètre astrométrique, ainsi que leurs unités de mesure :

1. RA (Right Ascension) : La RA est la coordonnée céleste qui mesure la position est-ouest des objets célestes, similaire à la longitude sur Terre. Elle est exprimée en heures, minutes et secondes d'arc (h, m, s).

2. DEC (Declination) : La DEC est la coordonnée céleste qui mesure la position nord-sud des objets célestes, similaire à la latitude sur Terre. Elle est également exprimée en degrés, minutes et secondes d'arc (°, ′, ″).

Le schéma ci-dessous illustre de manière claire les concepts de l'ascension droite (RA) et de la déclinaison (DEC). article graphicx

3. PRLX (Parallaxe) : La parallaxe est une mesure utilisée pour déterminer la distance des étoiles. Elle est exprimée en milli-arcsecondes (mas).

4. dPRLX (Incertitude de la Parallaxe) : L'incertitude de la parallaxe indique la marge d'erreur associée à la mesure de la parallaxe. Elle est exprimée en milli-arcsecondes (mas).

5. PM-RA (Mouvement propre en ascension droite) : Le mouvement propre en ascension droite mesure le déplacement apparent d'une étoile le long de l'ascension droite au fil du temps. Il est mesuré en milli-arcsecondes par an (mas/yr).

6. dPM-RA (Incertitude du mouvement propre en ascension droite) : incertitude du mouvement propre en ascension droite indique la précision de la mesure du mouvement propre d'une étoile le long de l'ascension droite. Elle est également exprimée en milli-arcsecondes par an (mas/yr).

7. PM-DEC (Mouvement propre en déclinaison) : Le mouvement propre en déclinaison mesure le déplacement apparent d'une étoile le long de la déclinaison au fil du temps. Il est mesuré en milli-arcsecondes par an (mas/yr).

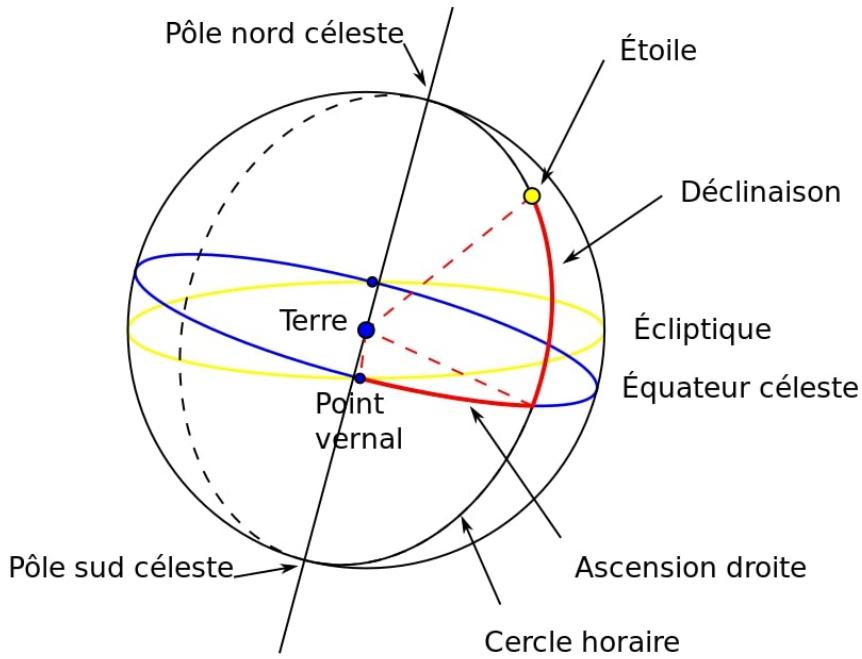


FIGURE 17 – Coordonnées équatoriales

8. dPM-DEC (Incertitude du mouvement propre en déclinaison) : L'incertitude du mouvement propre en déclinaison indique la précision de la mesure du mouvement propre d'une étoile le long de la déclinaison. Elle est également exprimée en milli-arcsecondes par an (mas/yr).

9. Julian Date : La date julienne est un système de mesure du temps utilisé en astronomie pour enregistrer le timing des observations astronomiques. Elle représente le nombre total de jours écoulés depuis une date de référence.

10. NS (Déplacement Nord-Sud) : Le déplacement nord-sud mesure le mouvement d'un objet céleste dans la direction nord-sud. Il est exprimé en milli-arcsecondes (mas).

11. DNS (Incertitude du déplacement Nord-Sud) : L'incertitude du déplacement nord-sud indique la précision de la mesure du mouvement d'un objet céleste dans la direction nord-sud. Elle est également exprimée en milli-arcsecondes (mas).

12. EW (Déplacement Est-Ouest) : Le déplacement est-ouest mesure le mouvement d'un objet céleste dans la direction est-ouest. Il est exprimé en milli-arcsecondes (mas).

13. dEW (Incertitude du déplacement Est-Ouest) : L'incertitude du déplacement est-ouest indique la précision de la mesure du mouvement d'un objet céleste dans la direction est-ouest. Elle est également exprimée en milli-arcsecondes (mas).

5.1.2 Importance en astrophysique

L'analyse du mouvement propre des étoiles joue un rôle crucial dans l'exploration de divers phénomènes astrophysiques et dans l'approfondissement de notre compréhension de l'univers. Ces études nous permettent de :

Cartographier la structure et la dynamique de la Voie lactée : En retracant les trajectoires des étoiles et en mesurant leurs vitesses, les astronomes peuvent reconstituer la distribution de matière dans notre galaxie et identifier les mouvements collectifs des étoiles qui révèlent sa structure et son évolution. Comprendre les processus de formation et d'évolution stellaire : L'étude du mouvement propre des étoiles jeunes permet de retracer leur origine dans les nuages de gaz et de poussière et d'évaluer leur vitesse d'éloignement de ces lieux de formation. De même, l'analyse du mouvement des étoiles plus âgées peut renseigner sur les processus d'interaction stellaire et d'évolution orbitale au sein de la galaxie. Détecter la présence d'exoplanètes : L'observation de légères perturbations dans le mouvement propre d'une étoile peut indiquer la présence d'une exoplanète en orbite autour d'elle. La mesure de ces perturbations permet d'estimer la masse et l'orbite de l'exoplanète, fournissant des informations précieuses sur la diversité des systèmes planétaires dans l'univers. Identifier les objets stellaires massifs et compacts : Les étoiles massives, telles que les étoiles à neutrons et les trous noirs, peuvent exercer une influence gravitationnelle importante sur leur environnement, perturbant le mouvement des étoiles voisines. L'analyse de ces perturbations permet d'identifier et de caractériser ces objets stellaires massifs et compacts, qui jouent un rôle crucial dans l'évolution

des galaxies. Étudier la formation et l'évolution des amas d'étoiles : En retracant les mouvements des étoiles au sein d'amas d'étoiles, les astronomes peuvent retracer l'histoire de formation de ces amas et identifier les interactions gravitationnelles qui façonnent leur structure et leur dynamique. En résumé, l'étude du mouvement propre des étoiles constitue un outil précieux pour les astrophysiciens, leur permettant d'explorer une vaste gamme de phénomènes cosmiques et de décrypter les mystères de l'univers. Ces études contribuent à enrichir notre compréhension de la formation et de l'évolution des étoiles, des galaxies et des systèmes planétaires, nous rapprochant ainsi d'une vision plus complète et plus précise de l'univers qui nous entoure.

5.2 Méthodologie

5.2.1 Data processing

Bases de données en astronomie, Simbad, JLP Horizons : Au cœur de l'Analyse Astrométrique réside la capacité à projeter avec précision les mouvements célestes dans le temps et l'espace. Notre objectif dans cette section est d'élaborer des algorithmes sophistiqués permettant de calculer les trajectoires temporelles des étoiles à partir des données fournies par les éphémérides JPL Horizons et les mesures de mouvement propre et de parallaxe de SIMBAD.

Avant d'aborder ces algorithmes, examinons de plus près ces sources de données. JPL Horizons fournit des éphémérides précises des corps célestes, telles que les positions et les vitesses des planètes, du Soleil et de la Lune. Ces données sont enregistrées sous forme de polynômes de Chebyshev ajustés aux coordonnées cartésiennes, couvrant un intervalle de temps spécifique et se référant à la version 2.0 de l'ICRF.

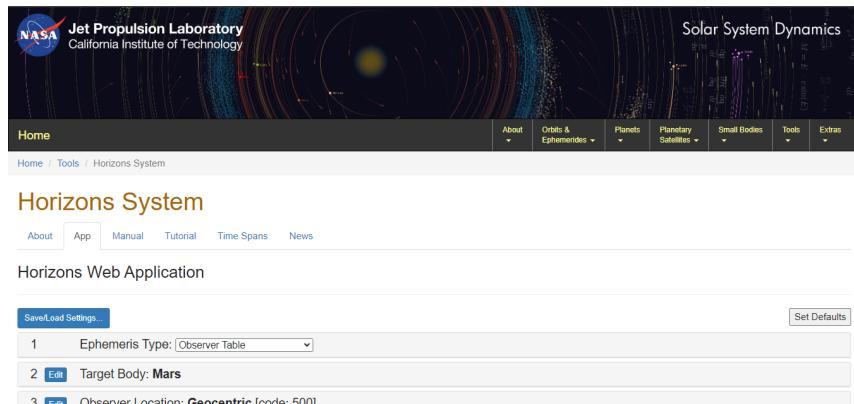


FIGURE 18 – JPL Horizons du Jet Propulsion Laboratory (JPL) de la NASA

D'autre part, SIMBAD constitue une ressource précieuse pour obtenir des informations sur les étoiles cibles utilisées dans notre analyse. Nous extrayons les données pertinentes telles que le nom, l'ascension droite (RA), la déclinaison (DEC), la parallaxe (PRLX), le mouvement propre dans les directions RA et DEC, ainsi que leurs erreurs, à partir de fichiers comme starlist.txt (ou CompendiumStars.txt).

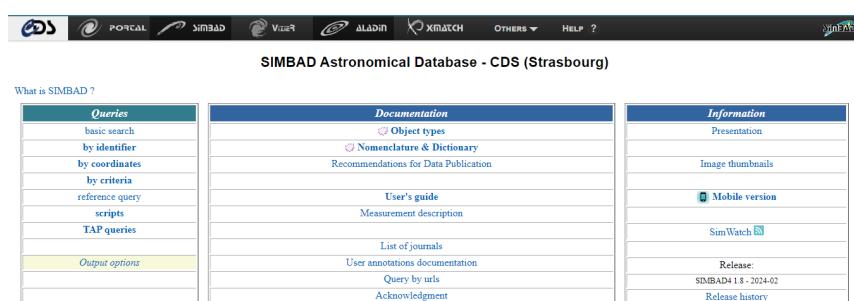


FIGURE 19 – La base de données SIMBAD

Création des fichiers nécessaires : Une fois que nous nous sommes familiarisés avec les sources de données astronomiques telles que SIMBAD et JPL Horizons, nous avons créé des fichiers de

données nécessaires à notre analyse.

- Le fichier "de430.bsp" renferme des informations sur les positions et les vitesses des planètes, du Soleil et de la Lune. Ces données sont enregistrées sous forme de polynômes de Chebyshev ajustés aux coordonnées cartésiennes, couvrant un intervalle de temps spécifique allant du 21 décembre 1549 au 25 janvier 2650, se référant à la version 2.0 de l'International Celestial Reference Frame (ICRF).
- Le fichier "starlist.txt" (ou "CompendiumStars.txt") agit comme une table de recherche fournissant des détails sur les étoiles cibles utilisées dans notre analyse. Il comprend des paramètres tels que le nom de l'étoile, son ascension droite (RA), sa déclinaison (DEC), sa parallaxe (PRLX), son mouvement propre dans les directions RA et DEC, ainsi que les erreurs associées à ces mesures.
Voici à quoi ressemble ce fichier :
HAT-P-7,19h28m59.3616s,+47d58m10.26s,+03.12, 0.44,-14.8,1.5,8.7,1.4
HAT-P-33,07h32m44.218s,+33d50m06.12s,+02.58, 0.06,-2.2,-4.9, 1.3, 1.0
HD129814,14h44m11.69s,+18d27m43.56s,24.32, 0.60,-70.88, 0.54,-151.63, 0.6
HD142229,15h53m20.01s,+04d15m11.50s,24.52, 1.25,-23.19, 1.08,+8.48, 1.24

- En outre, chaque dossier contient des mesures astrométriques spécifiques à une étoile donnée. Ces mesures sont fournies sous forme de Date Julianne, de décalage Nord-Sud (NS), de décalage Est-Ouest (EW), et de leurs erreurs respectives (dNS, dEW). Voici un exemple de ce fichier pour l'étoile HAT-P-7 :
2455803.8,-453.055,0.687,-110.458,0.714
2456522.9,-454.782,0.075,-111.798,0.201
2457213.3,-456.027,0.168,-77.298,0.811

5.2.2 Algorithmes de calcul

Le but du code est de modéliser le mouvement anticipé d'un objet céleste par rapport à une étoile cible donnée au fil du temps, en se basant sur diverses données astronomiques telles que la parallaxe, les mouvements propres et les observations astrométriques. Son objectif principal est de produire une trajectoire ou une courbe décrivant le déplacement prévu de l'objet céleste par rapport à l'étoile cible dans les directions Nord-Sud (NS) et Est-Ouest (EW), en fonction du temps. Cette trajectoire permet de prédire la position future de l'objet céleste, de planifier des observations astronomiques et d'étudier son mouvement relatif par rapport à une étoile de référence. De plus, le code vise à évaluer les erreurs associées à ces estimations pour garantir leur précision.

Voici un aperçu des principales étapes abordées par le code :

Importation des Modules

Le processus démarre par l'importation des bibliothèques essentielles au fonctionnement du code. Ces bibliothèques comprennent 'matplotlib', 'csv', 'jplephem', 'de421', 'numpy', 'astropy', 'datetime', et 'warnings'. En particulier, l'éphéméride DE430 ('de430.bsp') est ouverte à l'aide de 'jplephem'. Les constantes astrophysiques et les modules de gestion du temps sont également chargés. Des configurations initiales pour les visualisations avec 'matplotlib' sont également établies. Enfin, la date et l'heure de début d'exécution du code sont enregistrées pour référence ultérieure.

Définition des Fonctions

Au cœur du code se trouve une série de fonctions, dont la principale est `proj_RA_DEC`. Cette fonction est responsable du calcul des déplacements relatifs d'un objet céleste par rapport à une étoile de référence dans les directions NS (nord-sud) et EW (est-ouest). Voici un aperçu détaillé des différentes étapes de cette fonction :

1. **Calcul des déplacements dus au mouvement propre (proper motion) uniquement :**
 - Utilise la fonction `pm_offset` pour calculer les déplacements en RA et DEC dus au mouvement propre de l'objet céleste.
2. **Calcul des déplacements dus à la parallaxe (parallax) uniquement :**

- Utilise la fonction `parallax_offset` pour calculer les déplacements en RA et DEC dus à la parallaxe de l'objet céleste.
3. **Addition des déplacements :**
 - Ajoute les déplacements calculés précédemment en RA et DEC pour obtenir les déplacements totaux en RA et DE.
 4. **Obtention des positions initiales en RA et DEC :**
 - Utilise la fonction `get_initial_RA_DEC` pour calculer les positions initiales en RA et DEC de l'objet céleste.
 5. **Soustraction de la position initiale :**
 - Soustrait les positions initiales en RA et DEC des déplacements totaux en RA et DEC pour obtenir les changements nets en RA et DE.
 6. **Conversion en coordonnées NS (nord-sud) et EW (est-ouest) :**
 - Convertit les déplacements en RA et DEC en déplacements en NS et EW en utilisant les fonctions trigonométriques et la correction de la déclinaison pour l'EW.
 7. **Calcul des vecteurs NS et EW finaux :**
 - Ajoute les déplacements calculés aux déplacements initiaux en NS et EW pour obtenir les vecteurs NS et EW finaux.
 8. **Retour des vecteurs NS et EW :**
 - Retourne les vecteurs NS et EW, qui représentent les déplacements relatifs de l'objet céleste par rapport à l'étoile de référence.

En outre, d'autres fonctions auxiliaires sont utilisées pour exécuter cette fonction principale.

Paramètres et Configurations Initiales

Cette section du code est dédiée à la définition des paramètres et configurations initiales nécessaires à l'exécution des calculs ultérieurs.

1. **target :**
 - C'est une variable qui stocke le nom de la cible astronomique. Dans cet exemple, le nom est "HAT-P-7".
2. **params :**
 - C'est le nom du fichier qui contient les paramètres des étoiles. Dans cet exemple, il est nommé "starlist.txt".
3. **data_name :**
 - C'est le nom du fichier de données qui sera généré. Il semble être composé du nom de la cible (stocké dans `target`) suivi de "`_ast.txt`". Par exemple, si `target` est "HAT-P-7", alors `data_name` sera "HAT-P-7`_ast.txt`".
4. **npoints :**
 - C'est le nombre de points qui constituent le chemin en forme de "tornade". Il est défini ici à 100.
5. **ntracks :**
 - C'est le nombre de pistes. Dans le contexte de ce code, cela pourrait représenter le nombre de trajectoires ou de chemins calculés pour la cible spécifiée. Il est défini ici à 2000.

Lecture des Données et Conversion des Paramètres

Cette partie du code permet de traiter les données d'une étoile spécifique, de les convertir en radians, d'extraire et d'assigner les paramètres individuels, puis de calculer les positions NS et EW prédites sur une séquence de dates spécifiée. Voici les différentes opérations réalisées dans cette section :

1. **Lecture des données :**
 - Les données des étoiles sont lues à partir d'un fichier au format CSV (`params`) qui contient des informations telles que le nom de l'étoile, ses coordonnées (RA et DEC), sa parallaxe (PRLX), ses mouvements propres en RA et DEC (`PM_RA` et `PM_DEC`), ainsi que les incertitudes associées à ces paramètres.
 - Les données mesurées sont également lues à partir d'un autre fichier CSV (`data_name`) contenant les positions NS et EW ainsi que leurs incertitudes, enregistrées à différentes dates Julian.
2. **Conversion des paramètres en radians :**

- Les coordonnées de l'étoile (RA et DEC) sont converties en radians.
- Les autres paramètres (parallaxe, mouvements propres, incertitudes) sont convertis en radians.

3. Assiguation des paramètres individuels :

- Les valeurs des paramètres (parallaxe, mouvements propres, etc.) sont extraites des tableaux de données pour une étoile spécifique et assignées à des variables individuelles pour une manipulation plus aisée.

4. Calcul des positions NS et EW prédictes :

- Les positions NS et EW sont calculées pour une séquence de dates spécifiée (`t1`) en utilisant une fonction `proj_RA_DEC`.
- Cette fonction prend en compte les paramètres de l'étoile (RA, DEC, mouvements propres, parallaxe) ainsi que les dates spécifiées pour calculer les positions NS et EW prédictes.

Estimation des Erreurs

Dans cette section, nous explorons deux méthodes pour estimer les erreurs associées aux positions projetées des étoiles, suivies d'une discussion sur la visualisation des résultats pour évaluer la précision des prédictions et leur concordance avec les observations réelles.

1. Génération de Données Simulées

Nous débutons en générant des données simulées pour évaluer les erreurs associées aux positions estimées de l'étoile cible. Cette approche consiste à créer des données artificielles basées sur des distributions probabilistes centrées sur les valeurs mesurées, avec des écarts-types représentant les incertitudes des mesures.

- **Génération de Données Simulées** : Le code crée des données simulées à partir de distributions normales centrées sur les valeurs mesurées, avec des écarts-types égaux aux erreurs associées. Ces données imitent les mesures réelles et permettent d'évaluer les erreurs.
- **Calcul des Trajectoires Simulées** : Pour chaque ensemble de données simulées, la fonction `proj_RA_DEC` est utilisée pour calculer les trajectoires estimées de l'étoile dans les directions est et nord du ciel en fonction du temps. Ces trajectoires reflètent les différentes positions possibles de l'étoile en tenant compte de l'incertitude.
- **Détermination des Erreurs** : Les percentiles 16 et 84 des trajectoires simulées dans chaque direction sont calculés pour estimer l'incertitude sur la position de l'étoile dans les directions est et nord du ciel à différents moments.

2. Propagation des Erreurs

Nous utilisons ensuite la méthode de propagation des erreurs pour estimer l'incertitude dans les positions projetées des étoiles. Cette méthode prend en compte les incertitudes dans les mesures astrométriques et les paramètres astrométriques des étoiles.

- **Méthode de Propagation des Erreurs** : La propagation des erreurs est une méthode courante pour estimer l'incertitude dans une quantité calculée à partir de variables mesurées. Les équations de propagation des erreurs sont appliquées pour chaque composante des positions projetées (en NS et EW), en tenant compte des incertitudes associées.
- **Fiabilité des Positions Projetées** : Cette approche fournit une estimation quantitative de la fiabilité des positions projetées des étoiles, essentielle pour interpréter correctement les résultats de l'analyse astrométrique.
- **Exemple d'application dans l'étude** : Dans ce cas, la propagation des erreurs est réalisée en utilisant des équations complexes qui prennent en compte les différentes sources d'erreurs et les corrélations entre les paramètres. Ces équations sont dérivées en fonction des propriétés des mesures et des modèles utilisés. Dans l'exemple suivante (1) nous allons présenter l'équation de la variance de la composante de RA : l'ascension droite (σ_{ra}^2) :

$$\begin{aligned}
\sigma_{\text{ra}}^2 = & \sec^2(\text{DEC_J2000} + \text{pm_dec} \cdot \text{tl})^2 \times \\
& \left(\text{dprlx}^2 \cdot \left(y \cdot \cos(\text{RA_J2000} + \text{pm_ra} \cdot \text{tl} \cdot \sec(\text{DEC_J2000} + \text{pm_dec} \cdot \text{tl})) - \right. \right. \\
& \quad x \cdot \sin(\text{RA_J2000} + \text{pm_ra} \cdot \text{tl} \cdot \sec(\text{DEC_J2000} + \text{pm_dec} \cdot \text{tl})) \Big)^2 + \\
& \quad \text{dpm_ra}^2 \cdot \text{prlx}^2 \cdot \text{tl}^2 \cdot \sec^2(\text{DEC_J2000} + \text{pm_dec} \cdot \text{tl})^2 \cdot \\
& \quad \left(x \cdot \cos(\text{RA_J2000} + \text{pm_ra} \cdot \text{tl} \cdot \sec(\text{DEC_J2000} + \text{pm_dec} \cdot \text{tl})) + \right. \\
& \quad y \cdot \sin(\text{RA_J2000} + \text{pm_ra} \cdot \text{tl} \cdot \sec(\text{DEC_J2000} + \text{pm_dec} \cdot \text{tl})) \Big)^2 + \\
& \quad \text{dpm_dec}^2 \cdot \text{prlx}^2 \cdot \text{tl}^2 \cdot \\
& \quad \left(\cos(\text{RA_J2000} + \text{pm_ra} \cdot \text{tl} \cdot \sec(\text{DEC_J2000} + \text{pm_dec} \cdot \text{tl})) \cdot \right. \\
& \quad (-y + \text{pm_ra} \cdot \text{tl} \cdot x \cdot \sec(\text{DEC_J2000} + \text{pm_dec} \cdot \text{tl})) + \\
& \quad (x + \text{pm_ra} \cdot \text{tl} \cdot y \cdot \sec(\text{DEC_J2000} + \text{pm_dec} \cdot \text{tl})) \cdot \\
& \quad \sin(\text{RA_J2000} + \text{pm_ra} \cdot \text{tl} \cdot \sec(\text{DEC_J2000} + \text{pm_dec} \cdot \text{tl})) \Big)^2 \cdot \\
& \quad \tan^2(\text{DEC_J2000} + \text{pm_dec} \cdot \text{tl}) + \\
& \quad \text{prlx}^2 \cdot \sec^2(\text{DEC_J2000} + \text{pm_dec} \cdot \text{tl})^2 \cdot \\
& \quad \left(\text{dRA}^2 \cdot \left(x \cdot \cos(\text{RA_J2000} + \text{pm_ra} \cdot \text{tl} \cdot \sec(\text{DEC_J2000} + \text{pm_dec} \cdot \text{tl})) + \right. \right. \\
& \quad y \cdot \sin(\text{RA_J2000} + \text{pm_ra} \cdot \text{tl} \cdot \sec(\text{DEC_J2000} + \text{pm_dec} \cdot \text{tl})) \Big)^2 + \\
& \quad \text{dDEC}^2 \cdot \left(\cos(\text{RA_J2000} + \text{pm_ra} \cdot \text{tl} \cdot \sec(\text{DEC_J2000} + \text{pm_dec} \cdot \text{tl})) \cdot \right. \\
& \quad (-y + \text{pm_ra} \cdot \text{tl} \cdot x \cdot \sec(\text{DEC_J2000} + \text{pm_dec} \cdot \text{tl})) + \\
& \quad (x + \text{pm_ra} \cdot \text{tl} \cdot y \cdot \sec(\text{DEC_J2000} + \text{pm_dec} \cdot \text{tl})) \cdot \\
& \quad \sin(\text{RA_J2000} + \text{pm_ra} \cdot \text{tl} \cdot \sec(\text{DEC_J2000} + \text{pm_dec} \cdot \text{tl})) \Big)^2 \cdot \\
& \quad \tan^2(\text{DEC_J2000} + \text{pm_dec} \cdot \text{tl}) \Big) \quad (1)
\end{aligned}$$

Cette équation représente la variance de la composante de l'ascension droite (σ_{ra}^2) en fonction des différentes variables et de leurs incertitudes. En utilisant cette méthode, nous sommes en mesure d'évaluer de manière précise l'incertitude associée aux positions projetées des étoiles, ce qui est essentiel pour interpréter correctement les résultats de notre analyse astrométrique.

5.3 Visualisation des Résultats

Les résultats ont été présentés graphiquement pour une interprétation précise. Les décalages en déclinaison (NS) et en ascension droite (EW) ont été tracés pour fournir une représentation visuelle des variations spatiales des étoiles au fil du temps. De plus, la position angulaire (PA) et la séparation (SEP) ont été calculées pour une compréhension approfondie des changements dans les configurations stellaires. Le schéma ci-dessous illustre de manière claire les concepts de l'ascension droite (RA) et de la déclinaison (DEC).

5.3.1 Évolution du Mouvement Propre

Ci-dessus, la figure illustre les décalages NS et EW observés. La tendance descendante progressive du décalage NS suggère un déplacement de l'étoile vers le sud par rapport au référentiel céleste, tandis que la légère tendance à la hausse du décalage EW indique un déplacement vers l'ouest. Ces observations permettent d'estimer le mouvement propre de l'étoile dans les directions de la déclinaison et de l'ascension droite respectivement.

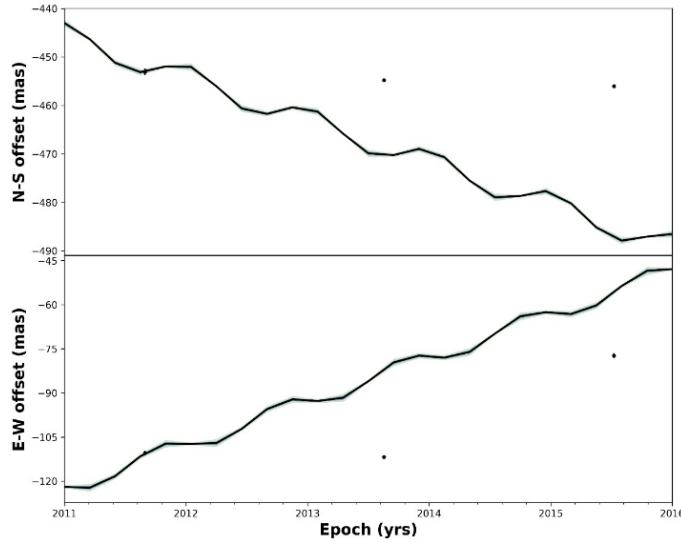


FIGURE 20 – Décalages NS et EW du mouvement propre de l'étoile

5.3.2 Évolution de la Position Angulaire (PA) et de la Séparation (SEP) de l'Étoile

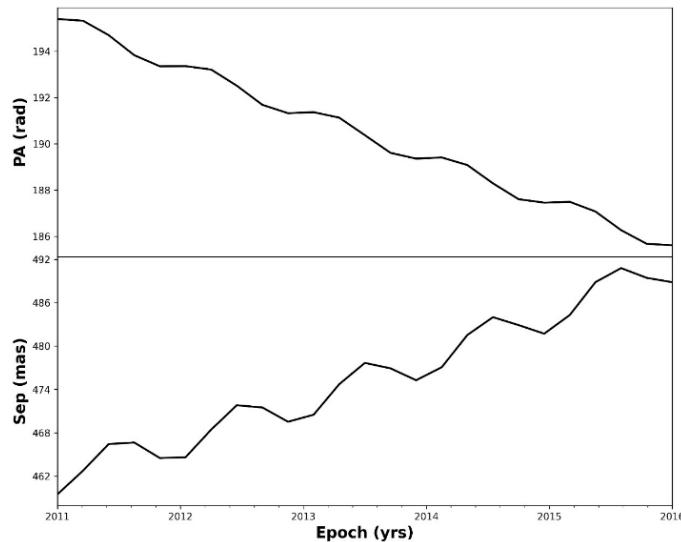


FIGURE 21 – Évolution de la Position Angulaire (PA) et de la Séparation (SEP) de l'Étoile

On observe que l'étoile suit la même tendance qu'auparavant, mais cette fois-ci avec la position angulaire (PA) et la séparation (SEP). La séparation de l'étoile semble augmenter au fil du temps, ce qui signifie que l'étoile s'éloigne de sa position de référence. De plus, la position angulaire augmente progressivement, suggérant que l'étoile pourrait suivre une trajectoire courbe par rapport à sa position initiale. Cette observation pourrait indiquer une orbite non circulaire dans un système binaire ou un mouvement propre complexe à travers le ciel.

5.3.3 Conclusion :

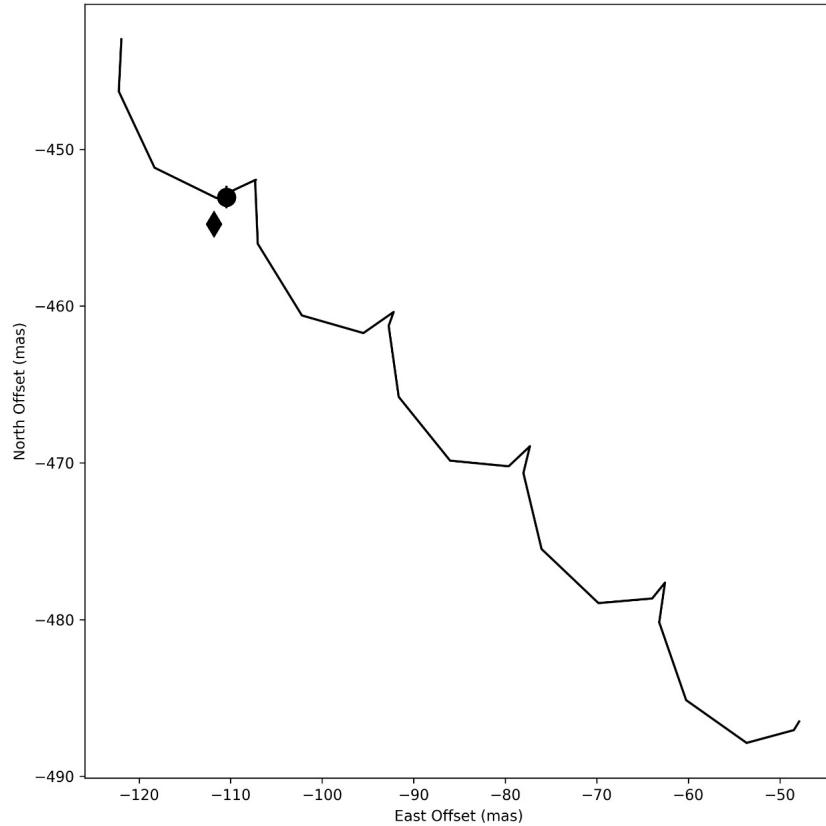


FIGURE 22 – Relation entre les décalages Nord et Est

L’analyse du mouvement propre de l’étoile révèle un comportement fascinant, caractérisé par une orientation constante vers le sud-ouest et une séparation croissante par rapport à sa position de référence(voir figure 8). Ce déplacement s’effectue selon une trajectoire cyclique, complétant une orbite approximative à intervalles réguliers. Au fil de ces cycles, l’étoile s’éloigne graduellement de sa position initiale, indiquant une évolution continue de son orbite.

Cette description met en lumière plusieurs aspects clés du mouvement de l’étoile :

- **Direction :** Le mouvement propre de l’étoile présente une composante sud-ouest dominante, indiquant une tendance à se déplacer dans cette direction.
- **Distance :** L’étoile s’éloigne continuellement de sa position de référence, suggérant une augmentation progressive de sa distance à ce point de repère.
- **Cyclicité :** Le mouvement de l’étoile suit une trajectoire cyclique, revenant périodiquement à proximité de sa position initiale avant de s’en éloigner à nouveau.
- **Évolution :** La séparation croissante de l’étoile par rapport à sa position de référence sur plusieurs cycles suggère une modification continue de son orbite, potentiellement due à des interactions gravitationnelles avec d’autres objets célestes.

6 Ouverture sur le Décryptage des Paramètres Morphologiques des Galaxies avec GaMPEN

6.1 Introduction

Il est important de noter que cette section ne couvre pas en détail la démonstration ou l'étude statistique approfondie de GaMPEN. Au lieu de cela, elle fournit une introduction générale à la solution et prépare le terrain pour une exploration plus approfondie. GaMPEN, acronyme de "Galaxy Morphology Posterior Estimation Network", est un modèle d'apprentissage profond révolutionnaire conçu pour estimer les paramètres morphologiques des galaxies avec une précision et une fiabilité inégalées. Contrairement aux approches traditionnelles qui prédisent des valeurs uniques pour ces paramètres, GaMPEN adopte une approche bayésienne sophistiquée, fournissant des distributions de probabilité complètes qui capturent l'incertitude inhérente à l'estimation.

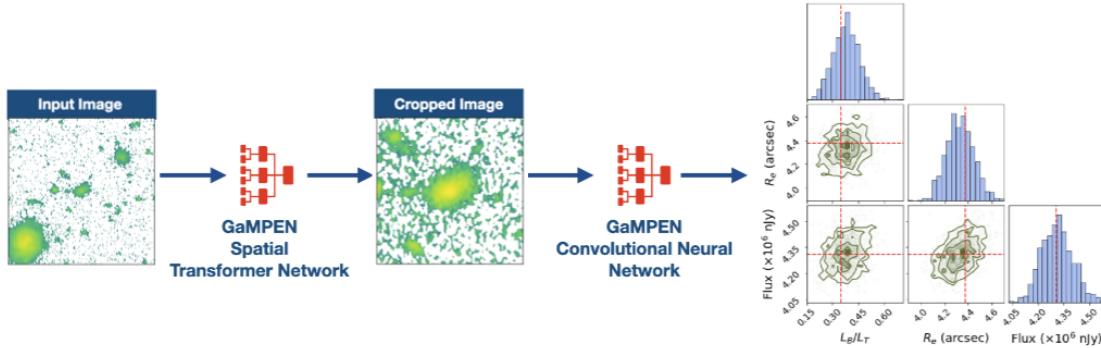


FIGURE 23 – gampen

GaMPEN s'attaque à l'estimation de trois paramètres morphologiques clés des galaxies :

(a) Rapport bulbe-disque (LB/LT)

Ce paramètre quantifie la proportion de lumière provenant du bulbe central de la galaxie par rapport à la lumière provenant du disque environnant. Un LB/LT élevé indique une galaxie dominée par un bulbe proéminent, tandis qu'un LB/LT faible suggère une galaxie dominée par un disque étendu.

Formule mathématique :

Le rapport bulbe-disque (LB/LT) est défini comme le rapport entre la luminosité totale du bulbe (LB) et la luminosité totale de la galaxie (LT).

$$\text{LB/LT} = \frac{\text{LB}}{\text{LT}}$$

(b) Rayon effectif (Re)

Le rayon effectif (Re) représente le rayon dans lequel la moitié de la lumière totale de la galaxie est contenue. Il s'agit d'une mesure de la taille de la galaxie.

Formule mathématique :

Le rayon effectif (Re) est généralement défini comme le rayon d'une sphère contenant la moitié de la luminosité totale de la galaxie.

$$\text{Re} = \sqrt{\frac{L}{\pi\Sigma}}$$

où :

L est la luminosité totale de la galaxie.

Σ est la luminosité de surface moyenne de la galaxie.

(c) Flux (F)

Le flux (F) représente la quantité totale de lumière émise par la galaxie dans une bande passante donnée. Il s'agit d'une mesure de la luminosité globale de la galaxie.

Formule mathématique :

Le flux (F) est généralement défini comme l'intégrale de la luminosité de surface de la galaxie sur toute sa surface.

$$F = \int \Sigma dA$$

où :

Σ est la luminosité de surface de la galaxie.

A est la surface de la galaxie.

L'estimation précise de ces paramètres morphologiques est cruciale pour comprendre l'évolution et la formation des galaxies. En effet, ces paramètres fournissent des informations précieuses sur la structure, la composition et l'histoire des galaxies. GaMPEN, en adoptant une approche bayésienne, offre un outil puissant pour estimer ces paramètres avec une précision et une fiabilité accrues, ouvrant ainsi la voie à une meilleure compréhension de l'univers des galaxies.

6.2 Implémentation Bayésienne de GaMPEN et Estimation des Incertitudes

Architecture de GaMPEN

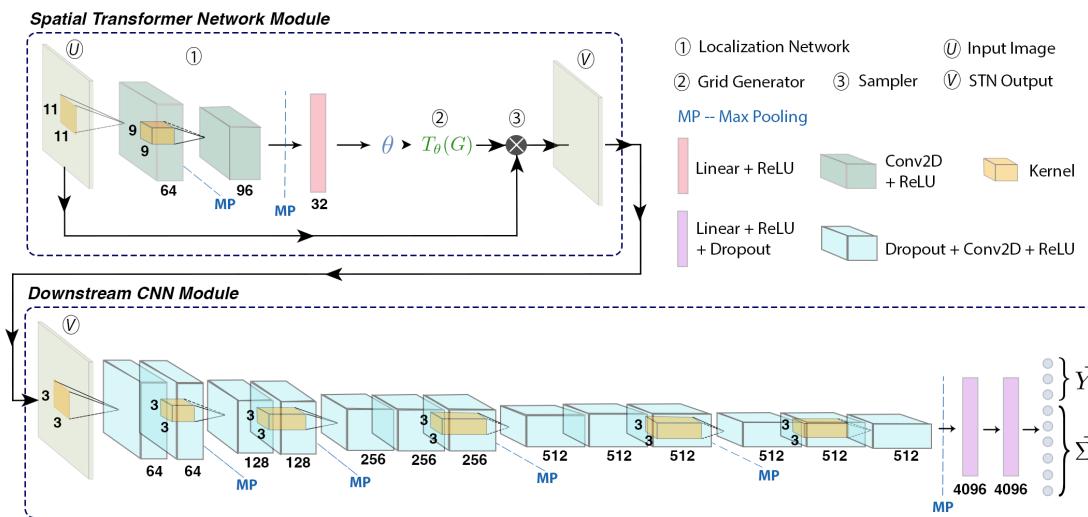


FIGURE 24 – GaMPEN se compose d'un réseau de transformation spatiale (STN) en amont suivi d'un réseau de neurones convolutionnel (CNN) en aval.

L'image fournie représente l'architecture du réseau d'apprentissage profond GaMPEN (Galaxy Morphology Posterior Estimation Network), conçu pour estimer les paramètres morphologiques des galaxies avec une précision et une fiabilité accrues. L'architecture se compose de deux modules principaux :

1. Module de transformateur spatial (STN)

Ce module joue un rôle crucial dans le prétraitement de l'image d'entrée. Il utilise une transformation affine bidimensionnelle pour recadrer automatiquement l'image, en se concentrant sur la galaxie elle-même et en éliminant les informations de fond non pertinentes. Cette étape préliminaire garantit que le CNN reçoit une image de la galaxie bien définie, améliorant ainsi la précision de l'analyse.

2. Réseau neuronal convolutif (CNN)

Le CNN constitue le cœur de GaMPEN, extrayant des caractéristiques discriminantes de l'image pré-traitée. Ces caractéristiques sont ensuite utilisées pour estimer les paramètres morphologiques de la galaxie. L'architecture du CNN peut impliquer plusieurs couches convolutives, suivies de couches entièrement connectées, pour aboutir aux distributions de probabilité finales des paramètres.

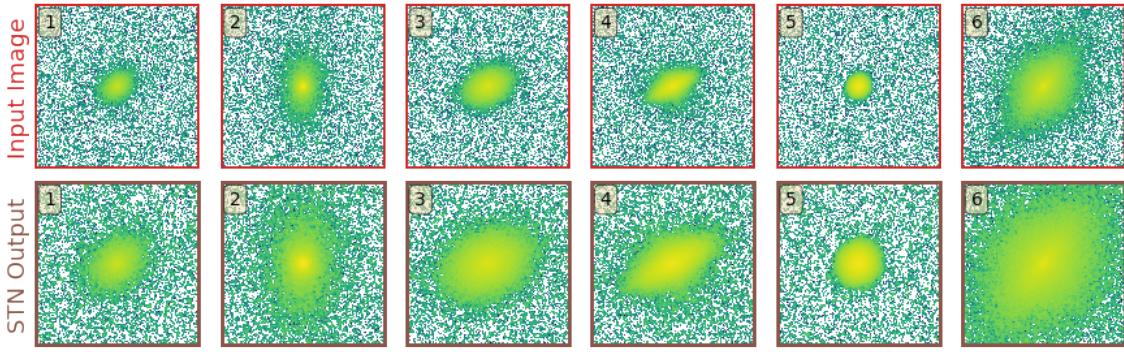


FIGURE 25 – Exemples de transformation appliquée par le STN à six images de galaxies simulées sélectionnées au hasard.

Explication du Module de Transformateur Spatial (STN)

Dans GaMPEN, le STN se situe en amont du CNN, où il applique une transformation affine bidimensionnelle à l'image d'entrée. Chaque image d'entrée est transformée différemment par le STN, qui apprend le recadrage approprié lors de l'entraînement du CNN en aval sans supervision supplémentaire. Pour effectuer la transformation, le STN prédit la matrice de transformation affine à six paramètres à appliquer à l'image d'entrée. Les coordonnées source et cibles sont modifiées par cette transformation affine, et le réseau apprend progressivement les transformations optimales en fonction de la fonction de perte.

Le placement du STN en amont dans le cadre de GaMPEN permet au réseau d'apprendre comment transformer activement l'image d'entrée de manière à minimiser la fonction de perte globale (c'est-à-dire que la transformation prédictive est guidée par et aide à l'estimation morphologique en aval). Cette propriété cruciale permet au STN d'être entraîné en utilisant la rétropropagation standard avec le CNN en aval, sans supervision supplémentaire.

Le principal avantage de l'utilisation d'un STN est que, sans besoin de formation supplémentaire, le STN apprend à recadrer systématiquement les galaxies secondaires dans les découpages et à se concentrer sur la galaxie d'intérêt au centre de la découpe.

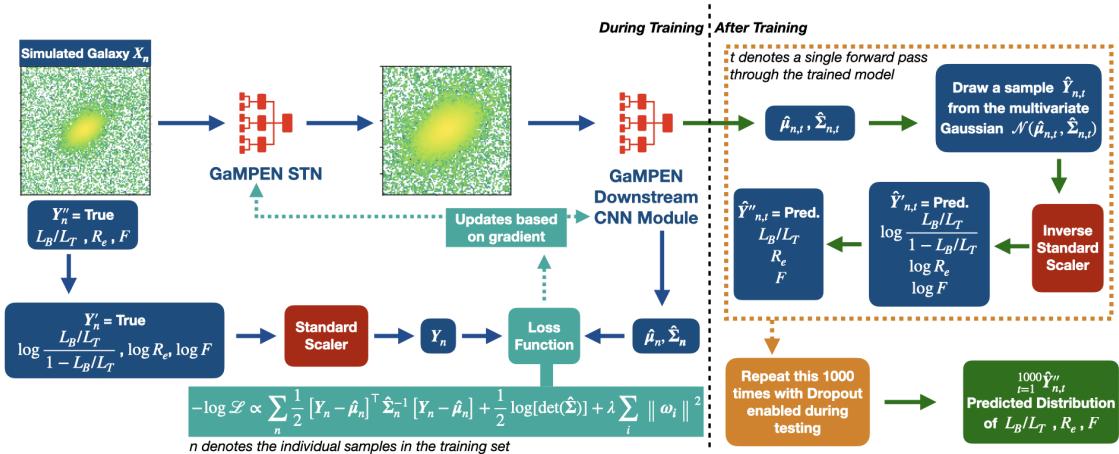


FIGURE 26 – Diagramme de flux de travail démontrant comment GaMPEN estime les distributions a posteriori

Estimation des Postérieurs/Prédiction des Incertitudes

Pour créer un cadre bayésien tout en prédisant les paramètres morphologiques, le modèle lui-même doit être traité comme une variable aléatoire - ou plus précisément, les poids du réseau doivent être des distributions probabilistes au lieu de valeurs uniques. Pour un réseau avec des poids ω , et un ensemble de données d'entraînement \mathcal{D} de taille N avec des images d'entrée $\{X_1, \dots, X_N\}$ et des paramètres de sortie $\{Y_1, \dots, Y_N\}$, le postérieur des poids du

réseau $p(\omega | \mathcal{D})$ représente les paramètres plausibles du réseau. Pour prédire la distribution de probabilité de la variable de sortie \hat{Y} donnée une nouvelle image de test \hat{X} , il faut marginaliser sur tous les poids possibles ω :

$$p(\hat{Y} | \hat{X}, \mathcal{D}) = \int p(\hat{Y} | \hat{X}, \omega) p(\omega | \mathcal{D}) d\omega.$$

Afin de calculer l'intégrale ci-dessus, il faut connaître $p(\omega | \mathcal{D})$, c'est-à-dire à quel point un ensemble particulier de poids est probable étant donné les données d'entraînement disponibles, \mathcal{D} . Puisque seul un modèle a été entraîné, il ne dit pas à quel point différents ensembles de poids sont probables. Différentes approximations ont été introduites afin de calculer cette distribution, avec l'inférence variationnelle étant la plus populaire.

La technique du dropout a été introduite par Srivastava et al. (2014) afin d'empêcher les réseaux neuronaux de surajuster ; ils ont temporairement supprimé des neurones aléatoires du réseau selon une distribution de Bernoulli, c'est-à-dire que des noeuds individuels étaient mis à zéro avec une probabilité, p , appelée taux de dropout. Ce processus de dropout peut également être interprété comme la permutation du modèle entraîné en un autre (Srivastava et al. 2014). En utilisant l'inférence variationnelle et le dropout, on peut approximer l'intégrale comme suit :

$$\int p(\hat{Y} | \hat{X}, \omega) p(\omega | \mathcal{D}) d\omega \approx \frac{1}{T} \sum_{t=1}^T p(\hat{Y} | \hat{X}, \omega_t)$$

où on effectue T passes avant avec le dropout activé et ω_t est l'ensemble des poids pendant la t -ème passe avant. Cette procédure est ce qu'on appelle le dropout Monte Carlo.

Pour obtenir des incertitudes épistémiques pour GaMPEN, une couche de dropout est insérée avant chaque couche de poids dans le réseau. Chaque passage avant à travers GaMPEN échantillonne le postérieur paramétrique approximatif. Ainsi, pour obtenir des incertitudes épistémiques, on alimente chaque image de test dans le cadre GaMPEN entraîné T fois et on collecte les sorties.

Calcul de la Vraisemblance et Incertitudes Aléatoires

L'ensemble de données d'entraînement se compose d'images d'entrée intentionnellement corrompues par du bruit, tandis que les paramètres morphologiques correspondants sont connus avec une précision parfaite. Cependant, en raison de quantités variables de bruit dans chaque image, les prédictions de GaMPEN au moment du test devraient présenter différents niveaux d'incertitudes. Ainsi, dans ce scénario, GaPEN est entraîné à prédire des incertitudes aléatoires.

Bien que GaMPEN soit destiné à prédire des incertitudes aléatoires, la matrice de covariance, Σ , n'est pas connue *a priori*. Au lieu de cela, GaMPEN est entraîné à apprendre ces valeurs en minimisant le logarithme négatif de la vraisemblance des paramètres de sortie pour l'ensemble d'entraînement, ce qui peut être écrit comme

$$-\log \mathcal{L}_{VI} \propto \sum_n \frac{1}{2} [\mathbf{Y}_n - \hat{\boldsymbol{\mu}}_n]^\top \hat{\boldsymbol{\Sigma}}_n^{-1} [\mathbf{Y}_n - \hat{\boldsymbol{\mu}}_n] + \frac{1}{2} \log [\det(\hat{\boldsymbol{\Sigma}}_n)] + \lambda \sum_i \|\omega_i\|^2$$

où $\hat{\boldsymbol{\mu}}^n$ et $\hat{\boldsymbol{\Sigma}}^n$ sont la moyenne et la matrice de covariance de la distribution gaussienne multivariée prédite par GaMPEN pour une image X_n . λ est la force du terme de régularisation, et ω_i sont échantillonnes à partir de $q(\omega)$. Notez que la fonction ci-dessus implique l'inverse et le déterminant de la matrice de covariance - son calcul peut être numériquement instable. Référez-vous au document GaMPEN pour savoir comment gérer cela.

Ainsi, pour prédire les incertitudes :

- (a) Pour chaque image, GaMPEN prédit les paramètres d'une distribution gaussienne multivariée (μ, Σ) . Ensuite, un échantillon est tiré de cette distribution.
- (b) Maintenant, le réseau est légèrement modifié en utilisant le dropout de Monte Carlo, et l'étape ci-dessus est répétée avec une estimation légèrement différente de (μ, Σ) .
- (c) La dernière étape est maintenant répétée 1000 fois pour chaque galaxie.

La combinaison des deux étapes ci-dessus nous permet d'estimer des incertitudes robustes.

6.3 Test de la Solution

Dans cette section, nous avons réalisé des tests de la solution en utilisant des modèles GaMPEN pré-entraînés pour prédire des caractéristiques clés à partir d'images de galaxies. Le processus de test a été effectué en suivant les étapes détaillées ci-dessous :

- (a) **Préparation de l'environnement de test** : Nous avons configuré l'environnement de test en installant GaMPEN et en téléchargeant les images de galaxies ainsi que les modèles pré-entraînés.
- (b) **Inference sur les images de galaxies** : En utilisant les modèles GaMPEN pré-entraînés, nous avons effectué des prédictions sur les images de galaxies à différentes plages de décalage vers le rouge (z).
- (c) **Analyse des résultats** : Les prédictions obtenues ont été analysées pour évaluer la précision et la fiabilité de la solution. Nous avons calculé des statistiques sommaires telles que les moyennes, médianes, écarts-types, ainsi que des intervalles de confiance pour évaluer la distribution des prédictions.
- (d) **Visualisation des résultats** : Les résultats ont été visuellement inspectés en comparant les distributions prédictives avec les images originales des galaxies. Des graphiques ont été générés pour représenter les distributions prédictives des caractéristiques clés telles que le rapport luminosité/taille (L_B/L_T), le rayon effectif (R_e), et la magnitude.
- (e) **Evaluation des performances** : Les performances de la solution ont été évaluées en tenant compte de la capacité des modèles à fournir des prédictions cohérentes et précises par rapport aux caractéristiques réelles des galaxies.

6.3.1 Entrées et Sorties

Dans cette sous-section, nous détaillons les différentes entrées et sorties du processus de test. Cela inclut les données utilisées pour l'inférence ainsi que les résultats obtenus à partir de cette inférence.

Entrées Les entrées du processus de test comprennent les éléments suivants :

- **Images de Galaxies** : Nous avons utilisé des images de galaxies provenant de différentes sources, telles que le Hyper Suprime Cam Imaging, pour effectuer les prédictions. Un exemple d'image de galaxie est illustré à la Figure 27.
- **Modèles Pré-Entraînés** : Les modèles GaMPEN pré-entraînés ont été utilisés pour réaliser l'inférence sur les images de galaxies. Ces modèles ont été téléchargés à partir de sources fiables. Un exemple de modèle pré-entraîné est illustré à la Figure 28.

Sorties Les sorties du processus de test comprennent les éléments suivants :

- **Prédictions des Caractéristiques** : Les modèles GaMPEN ont généré des prédictions pour différentes caractéristiques des galaxies, telles que le rapport luminosité/taille (L_B/L_T), le rayon effectif (R_e), et la magnitude. Un exemple de prédictions est illustré à la Figure 29.
- **Statistiques Sommaires** : Des statistiques sommaires ont été calculées à partir des prédictions pour évaluer la distribution et la précision des résultats.
- **Graphiques de Visualisation** : Des graphiques ont été générés pour représenter visuellement les distributions prédictives des caractéristiques des galaxies.

Ces entrées et sorties ont été utilisées pour évaluer la performance globale de la solution et déterminer sa fiabilité pour des applications futures.

7 Bibliographie

- Patrick Hébert Denis Laurendeau, Traitement des images , Université LAVAL.
- Maïtine Bergounioux. Quelques méthodes de filtrage en Traitement d'Image. 2010.
hal-00512280v1
- Bryan S. Morse, Thresholding, Brigham Young University, 1998–2000
- Stephens, M. A. (1974). EDF Statistics for Goodness of Fit and Some Comparisons, Journal of the American Statistical Association, 69, pp. 730-737.
- Stephens, M. A. (1976). Asymptotic Results for Goodness-of-Fit Statistics with Unknown Parameters, Annals of Statistics, 4, pp. 357-369.
- Stephens, M. A. (1977). Goodness of Fit for the Extreme Value Distribution, Biometrika, 64, pp. 583-588.
- Stephens, M. A. (1977). Goodness of Fit with Special Reference to Tests for Exponentiality , Technical Report No. 262, Department of Statistics, Stanford University, Stanford, CA.
- [James Webb \(Image analysis\), Kaggle](#)
- [Astrophysical-Image-Processing-Using-JWST, Github](#)
- [Otsu method, Wikipedia](#)
- [Canny Detection, OpenCV documentation.](#)
- [Canny Edge Detection Step by Step in Python — Computer Vision, towardsdatascience.com](#)

8 Annexe

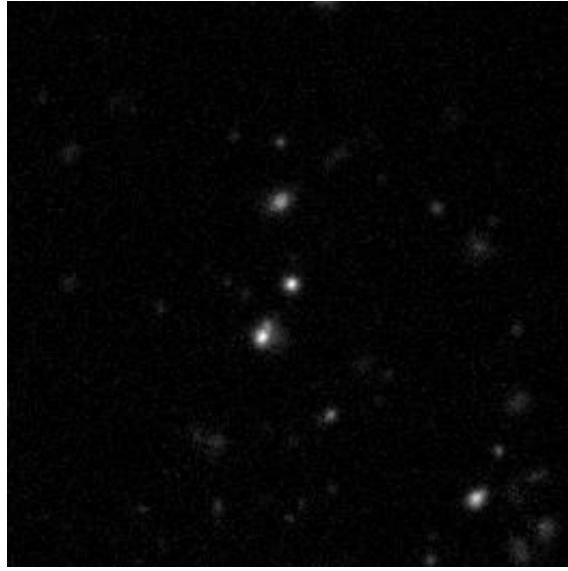


FIGURE 27 – Exemple d’image de galaxie utilisée comme entrée pour l’inférence.

```
[ ] %cd /content/  
!python GaMPEN/ggt/modules/inference.py  
--model_path='trained_models/g_0_025_real_data.pt'\  
--output_path="bayesian_inference_runs_lowz/"\  
--data_dir='data_lowz/'\  
--split=None\  
--slug=None\  
--cutout_size=239\  
--normalize\  
--transform\  
--n_workers=2\  
--parallel\  
--label_cols='custom_logit_bt,ln_R_e_asec,ln_total_flux_adus'\  
--model_type='vgg16_w_stn_oc_drp'\  
--repeat_dims\  
--channels=3\  
--label_scaling='std'\  
--mc_dropout\  
--cov_errors\  
--dropout_rate=0.0004\  
--n_runs=50\  
--no_labels\  
--scaling_data_dir='scaling_data_dir_lowz/'\  
--scaling_slug='balanced-dev2'
```

FIGURE 28 – Exemple de modèle pré-entraîné utilisé pour réaliser l’inférence sur les images de galaxies.

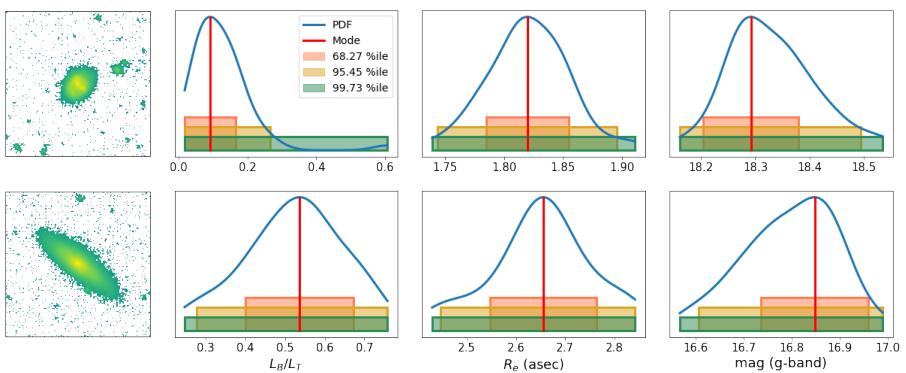


FIGURE 29 – Exemple de prédictions générées par les modèles GaMPEN pour les caractéristiques des galaxies.