

Projet de Fin d'Études

"Système de Suivi et d'Analyse des Émotions d'un Bébé en Temps Réel"

Ce projet consiste à développer un système capable de détecter et suivre les émotions d'un bébé en temps réel à partir de son expression faciale, afin d'aider les parents à mieux comprendre ses besoins émotionnels.

Réalisé par : Jaballah Mootaz

Filière : Intelligence Artificielle et Data Science

Année Universitaire : 2024-2025

EPI DIGITAL SCHOOL

Encadré par (mdm) : Jalouli Malika

Remerciements

Je tiens à remercier toutes les personnes qui ont contribué au succès de mon stage et qui m'ont aidé lors de la rédaction de ce rapport. Je voudrais dans un premier temps adresser mes sincères remerciements à M. Cheheb Achref de m'avoir orienté et encouragé à réaliser ce travail ainsi que pour la confiance qu'il m'a accordée en acceptant de m'encadrer dans cette société. aussi pour Madame Malika Jalouli pour son encadrement pédagogique et ses conseils plus qu'utiles J'aimerais également le remercier pour l'autonomie qu'il m'a laissée, ainsi que pour ses précieux conseils qui m'ont permis de mener à bien ce travail. Je remercie aussi tout le corps enseignant et la direction de l'EPI pour leur dévouement dans notre formation et leur encadrement tout au long de cette année. Merci à tous.

Table des matières

Remerciements	1
Introduction	3
1 État de l'art	6
1.1 <i>Les émotions chez les nourrissons</i>	6
1.2 <i>Vision par ordinateur et reconnaissance des émotions</i>	7
1.3 <i>Apprentissage profond (Deep Learning) et CNN</i>	8
1.4 <i>Travaux connexes et études similaires</i>	9
1.5 <i>Comparaison des approches existantes</i>	9
1.6 <i>Justification du choix technologique</i>	10
2 Description du projet	12
2.1 <i>Présentation générale du système</i>	12
2.2 <i>Spécifications fonctionnelles</i>	12
2.3 <i>Spécifications non fonctionnelles</i>	13
2.4 <i>Architecture fonctionnelle</i>	13
2.5 <i>Diagramme de cas d'utilisation</i>	13
2.6 <i>Environnement de développement</i>	14
2.7 <i>Structure technique de l'application</i>	15
2.8 <i>Gestion du dataset et annotation</i>	16

Introduction

Contexte Général

Dans une société où l'intelligence artificielle (IA) s'intègre progressivement dans les domaines de la santé, de l'éducation et du bien-être, les besoins en systèmes intelligents de surveillance et d'assistance deviennent de plus en plus pressants. L'un des groupes les plus vulnérables et dépendants est celui des nourrissons. Ces derniers, ne pouvant s'exprimer verbalement, utilisent principalement leurs expressions faciales et leurs pleurs pour communiquer leurs besoins et émotions.

Aujourd'hui, il devient techniquement envisageable de traduire ces signaux visuels en émotions interprétables grâce aux progrès du traitement d'image et des réseaux neuronaux convolutifs. Cela ouvre la voie à des applications d'assistance en temps réel permettant d'interpréter l'état émotionnel d'un bébé pour mieux répondre à ses besoins.

Problématique

Les approches classiques d'analyse émotionnelle chez les nourrissons reposent principalement sur l'observation humaine ou sur l'usage de capteurs physiologiques (fréquence cardiaque, température, sudation, etc.). Bien que pertinentes, ces méthodes se heurtent à plusieurs limites : elles sont souvent coûteuses, intrusives et peu accessibles aux parents dans un contexte domestique.

Dans notre cas, l'absence de matériel IoT nous a conduit à explorer une alternative plus pratique et économique : l'analyse des émotions à partir de vidéos en temps réel, sans contact physique avec l'enfant, à l'aide de la vision par ordinateur. Cette approche vise à exploiter uniquement la caméra d'un téléphone portable pour détecter et classifier les émotions du bébé.

Objectifs du Projet

Le présent projet de fin d'études a pour but de concevoir un système intelligent de suivi et d'analyse des émotions d'un bébé en temps réel, basé sur une architecture d'IA déployée dans une application mobile développée avec Flutter. Plus précisément, ce projet vise à :

- Mettre en œuvre un modèle d'intelligence artificielle pour la reconnaissance faciale des émotions.
- Intégrer ce modèle (Yolo v11) dans une application mobile Flutter optimisée pour les smartphones Android.
- Utiliser des vidéos de visages de nourrissons comme source d'entrée, sans recours à des capteurs biométriques.
- Faciliter l'analyse par l'utilisateur (parents, médecins, puériculteurs) via une interface conviviale.
- Fournir des alertes ou des rapports de suivi en fonction des émotions détectées.

Méthodologie Adoptée

La méthodologie suivie dans ce projet repose sur une approche expérimentale et itérative. Les principales étapes sont :

1. **Étude bibliographique** : une revue approfondie des travaux existants sur la reconnaissance des émotions, les modèles de deep learning, et les outils de détection faciale.
2. **Collecte et annotation des données** : en l'absence de base de données spécifique aux bébés, nous avons utilisé la plateforme Roboflow pour créer et annoter un dataset d'images faciales.
3. **Choix et entraînement du modèle** : nous avons adopté le modèle Yolo v11, réputé pour sa performance en temps réel, en l'adaptant à notre contexte.
4. **Développement de l'application** : implémentation d'une application mobile en Flutter, intégration du modèle, conception de l'interface utilisateur.
5. **Évaluation et validation** : tests sur divers scénarios, évaluation des performances, ajustement des paramètres.

Structure du Rapport

Ce document est structuré comme suit :

- **Chapitre 1** : Présentation des concepts de base et des approches existantes dans la littérature.
- **Chapitre 2** : Description complète du système à concevoir, incluant les spécifications et les outils utilisés.
- **Chapitre 3** : Architecture technique et implémentation du modèle IA et de l'application mobile.
- **Chapitre 4** : Résultats expérimentaux, tests et interprétation des performances.
- **Chapitre 5** : Déploiement du modèle dans un environnement mobile et perspectives d'industrialisation.
- **Conclusion** : Synthèse des travaux réalisés, limites rencontrées et perspectives d'amélioration.

Chapitre 1

État de l’art

1.1 Les émotions chez les nourrissons

Le développement émotionnel des nourrissons est un domaine complexe mêlant psychologie, neurosciences et sciences du comportement. Dès les premières semaines de vie, les bébés expriment des émotions fondamentales telles que la joie, la peur, la tristesse et la colère. Ces émotions ne sont pas seulement visibles dans leurs pleurs ou leurs rires, mais aussi à travers des micro-expressions faciales, souvent imperceptibles à l’œil non averti.

Selon les travaux de Paul Ekman, certaines expressions faciales sont universelles, c’est-à-dire qu’elles sont présentes dans toutes les cultures humaines dès le plus jeune âge. Chez le nourrisson, on observe notamment :

- Le sourire social dès 6 à 8 semaines.
- La peur manifeste à partir de 6-7 mois.
- Des signes de frustration ou de colère lorsqu’un besoin n’est pas satisfait.

Ce développement suit un schéma progressif :

- **0–3 mois** : réponses réflexes, émergence du plaisir simple (contact, chaleur).
- **3–6 mois** : premières différenciations émotionnelles, début des sourires intentionnels.
- **6–12 mois** : peur de l’étranger, anxiété de séparation, communication affective par gestes et regards.

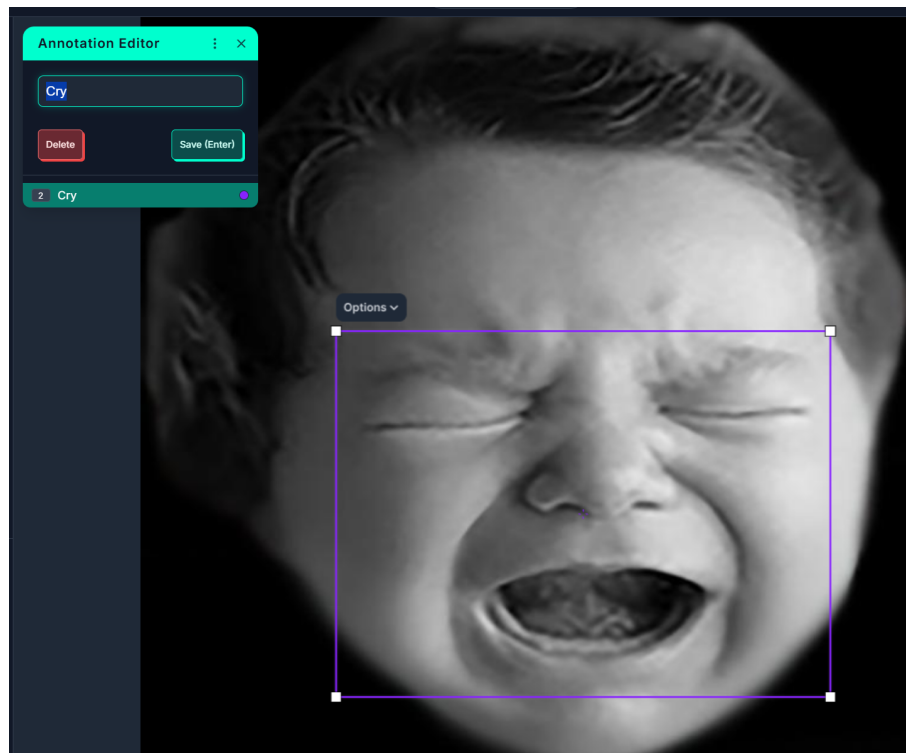


FIGURE 1.1 – Exemples d'expressions faciales typiques chez les bébés.

L'identification automatisée de ces émotions permettrait aux parents ou soignants de mieux répondre aux besoins affectifs du bébé, surtout en leur absence directe (garderie, nuit, etc.).

1.2 Vision par ordinateur et reconnaissance des émotions

La vision par ordinateur (Computer Vision) est une branche de l'intelligence artificielle qui vise à permettre à une machine de "voir" et d'interpréter des images de manière automatisée. Lorsqu'elle est appliquée à la reconnaissance des émotions, elle cherche à détecter les changements subtils dans les traits du visage, comme :

- la courbure des lèvres (sourire, moue),
- l'élévation des sourcils (surprise, détresse),
- le plissement des yeux (colère, douleur),
- la tension musculaire autour de la bouche ou du menton.

Ce processus passe généralement par trois étapes clés :

1. **Détection du visage** : localisation du visage dans une image ou une séquence vidéo.

2. **Extraction de caractéristiques** : calcul de descripteurs visuels (landmarks, textures, gradients...).
3. **Classification émotionnelle** : attribution d'une étiquette émotionnelle à partir des patterns détectés.

Les systèmes modernes s'appuient de plus en plus sur l'apprentissage profond pour automatiser ces étapes sans ingénierie manuelle des caractéristiques.

1.3 Apprentissage profond (Deep Learning) et CNN

L'apprentissage profond (deep learning) est un sous-domaine du machine learning reposant sur des réseaux de neurones artificiels à multiples couches. Ces réseaux, lorsqu'ils sont convolutifs (CNN), sont particulièrement adaptés aux tâches d'analyse d'images.

Un CNN est structuré en :

- **Couches de convolution** : détectent les motifs visuels locaux.
- **Couches de pooling** : réduisent la dimensionnalité tout en conservant l'information essentielle.
- **Couches entièrement connectées** : prennent les décisions finales (ici : classification émotionnelle).

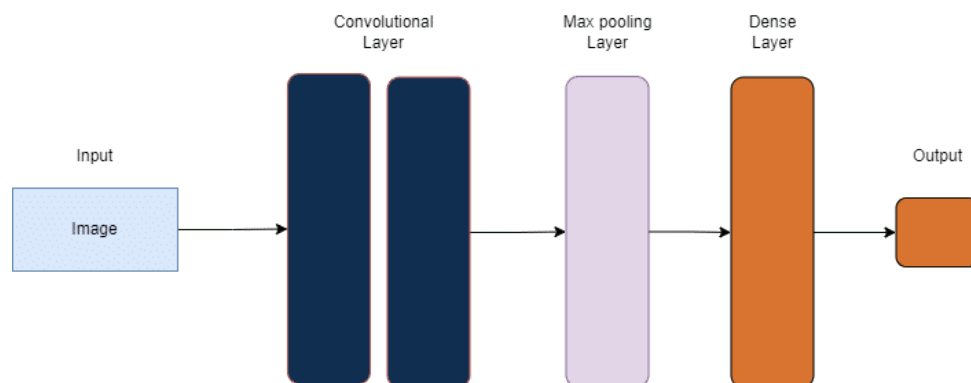


FIGURE 1.2 – Architecture simplifiée d'un réseau de neurones convolutif (CNN).

Les CNN offrent une excellente robustesse aux variations de luminosité, de position du visage, ou de légères occlusions (jouets, tétine...). Pour notre projet, l'architecture YOLO (You Only Look Once) a été préférée pour sa capacité à effectuer la détection et la classification simultanément et en temps réel.

1.4 Travaux connexes et études similaires

De nombreux chercheurs ont exploré la détection des émotions à partir des images de visages, mais très peu l'ont fait spécifiquement pour les nourrissons. Voici quelques exemples représentatifs :

- **Zhao et al. (2020)** : ont utilisé une caméra thermique combinée à un CNN pour détecter les émotions liées à l'anxiété chez les enfants.
- **DeepFaceEmotion** : plateforme s'appuyant sur VGG16 pour reconnaître sept émotions humaines standard (joie, peur, colère, etc.) chez l'adulte.
- **BabyEmotionNet (2022)** : propose une adaptation de ResNet50 entraînée sur un dataset artificiel de visages de nourrissons générés par GANs.

Ces approches restent toutefois limitées :

- Utilisation d'équipements non accessibles au grand public.
- Faible nombre de données réelles de visages de nourrissons annotées émotionnellement.
- Temps d'inférence élevé, peu compatible avec un usage sur mobile.

1.5 Comparaison des approches existantes

TABLE 1.1 – Comparaison entre différentes approches de détection émotionnelle

Approche	Technologie utilisée	Avantages / Inconvénients
DeepFaceEmotion	CNN (VGG16)	Reconnaissance émotionnelle adulte, pré-entraîné, mais peu adaptable aux enfants
BabyEmotionNet	ResNet50 sur dataset synthétique	Adapté aux bébés, mais dataset peu fiable et manque de diversité réelle
Notre approche	Yolo v11 + annotation Roboflow + Flutter	Temps réel, léger, fonctionne sans capteur, adapté à une application mobile

1.6 Justification du choix technologique

Le choix de YOLO v11 repose sur plusieurs critères techniques et stratégiques :

- **Vitesse d'exécution** : YOLO traite une image en un seul passage (single shot), ce qui permet une détection en temps réel.
- **Compatibilité mobile** : possibilité de convertir le modèle vers ONNX puis TFLite pour l'intégration directe dans une application Flutter.
- **Taille optimisée** : contrairement à ResNet ou VGG, les versions récentes de YOLO sont plus compactes tout en maintenant une bonne précision.

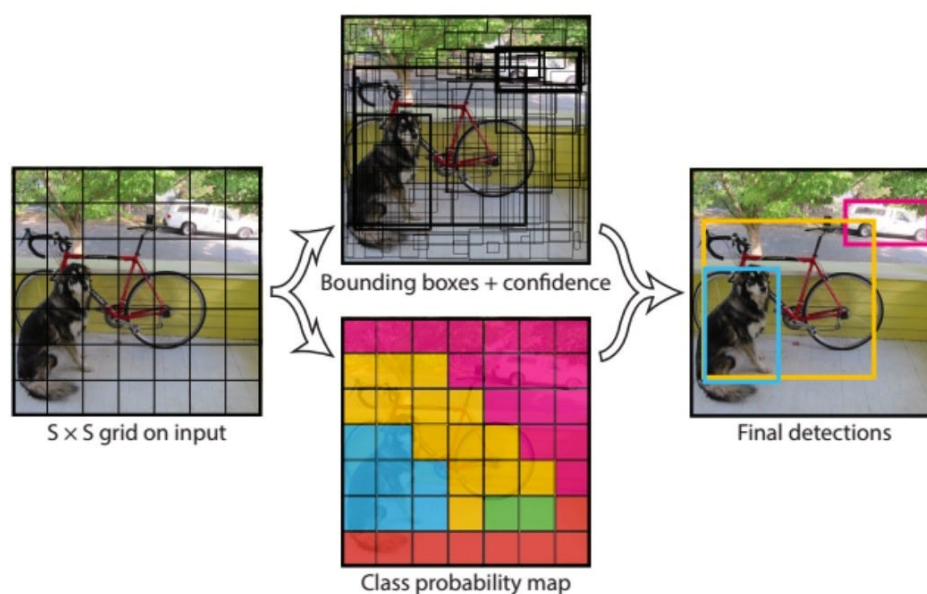


FIGURE 1.3 – Principe de fonctionnement de l'architecture YOLO (You Only Look Once).

En parallèle, Flutter est un framework moderne permettant de développer des interfaces utilisateur performantes, multi-plateformes et élégantes. Il offre également une bonne intégration avec les bibliothèques d'intelligence artificielle, notamment via les plugins 'tflite_flutter' ou 'google_ml_kit'.

Conclusion partielle

Ce chapitre a permis de poser les fondations scientifiques du projet. Il montre que l'analyse des émotions chez les nourrissons représente un défi interdisciplinaire à la croisée de la psychologie, de la vision artificielle et du développement mobile. Les approches existantes, bien qu'intéressantes, manquent souvent de souplesse ou de réalisme. Notre projet

ambitionne donc de combiner les meilleures pratiques de l'intelligence artificielle temps réel avec la simplicité d'un usage quotidien sur smartphone.

Chapitre 2

Description du projet

2.1 Présentation générale du système

Le système que nous avons conçu est une application mobile intelligente destinée à surveiller et interpréter les émotions des nourrissons en temps réel. Il s'agit d'un outil d'assistance à la parentalité, basé sur des technologies d'intelligence artificielle intégrées directement dans un smartphone.

Fonctionnement global : à l'aide de la caméra du téléphone, l'application capture des images du visage du bébé, les analyse à l'aide d'un modèle pré-entraîné basé sur YOLO v11, puis affiche l'émotion détectée (ex. : joie, tristesse, détresse) à l'utilisateur, tout en sauvegardant un historique des états émotionnels dans un journal local.

Ce système est conçu pour être non intrusif, autonome (sans besoin d'internet), et ergonomique pour un usage quotidien.

2.2 Spécifications fonctionnelles

Les spécifications fonctionnelles décrivent les fonctionnalités attendues du système du point de vue de l'utilisateur final.

- L'utilisateur peut lancer la détection des émotions en un seul clic.
- L'application affiche en direct l'émotion détectée.
- Un journal émotionnel est généré et consultable sous forme de graphique ou de tableau.
- L'application permet la gestion multi-bébés (plusieurs profils).
- Aucune connexion internet n'est requise pour la détection.

2.3 Spécifications non fonctionnelles

- **Performance** : l'inférence du modèle IA doit être réalisée en moins de 300 ms.
- **Compatibilité** : application mobile Android à partir d'Android 8.0.
- **Confidentialité** : aucune donnée n'est transmise à un serveur externe.
- **Scalabilité** : possibilité d'intégration future avec des objets connectés (babycam, smartwatch...).

2.4 Architecture fonctionnelle

Le système peut être divisé en trois modules principaux :

1. **Module de capture** : permet de récupérer les images de la caméra.
2. **Module d'analyse IA** : détecte le visage du bébé et déduit son émotion.
3. **Module d'interface** : affiche les résultats à l'utilisateur et gère les historiques.

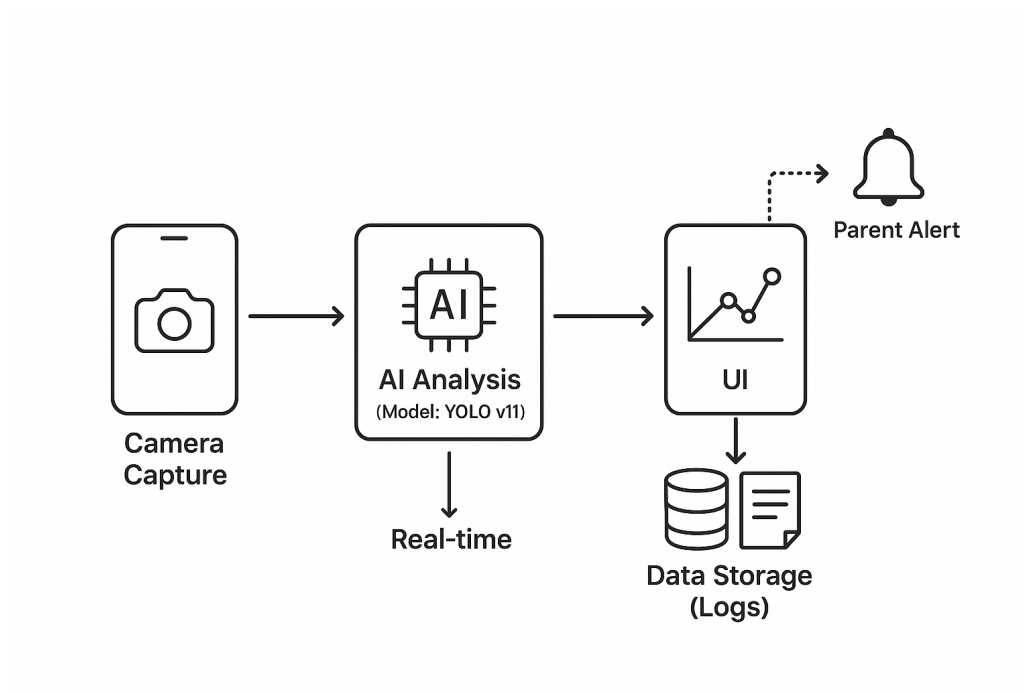


FIGURE 2.1 – Architecture fonctionnelle générale du système.

2.5 Diagramme de cas d'utilisation

Ce diagramme décrit les interactions principales entre l'utilisateur et le système.



FIGURE 2.2 – Diagramme de cas d'utilisation principal.

2.6 Environnement de développement

Langages et frameworks

- **Flutter** : framework UI pour développer l'application mobile multiplateforme.
- **Dart** : langage principal utilisé avec Flutter.
- **Python** : utilisé pour entraîner, convertir, ou tester le modèle YOLO v11.

Outils utilisés

- **Roboflow** : outil de gestion, annotation et génération de dataset, avec export compatible YOLO.

- **Google Colab** : pour l'entraînement du modèle en GPU.
- **ONNX/TFLite** : pour convertir le modèle IA afin de l'intégrer dans Flutter.
- **Android Studio** : pour compiler, tester et déployer l'application.

2.7 Structure technique de l'application

L'application Flutter a été structurée selon une architecture en couches :

- **UI Layer** : écrans, widgets, affichage des émotions.
- **Logic Layer** : gestion des caméras, appels aux modèles IA, gestion des utilisateurs.
- **Model Layer** : chargement et exécution du modèle YOLO v11.



FIGURE 2.3 – Architecture logicielle de l'application Flutter.

2.8 Gestion du dataset et annotation

La base de données utilisée a été préparée manuellement à l'aide de la plateforme Roboflow. Voici le processus adopté :

1. Collecte d'images représentatives d'émotions faciales infantiles.
2. Annotation via Roboflow : chaque visage est étiqueté selon son émotion apparente.
3. Génération automatique des fichiers d'entraînement YOLO (images + fichiers .txt).
4. Division du dataset : 80% entraînement, 10% validation, 10% test.

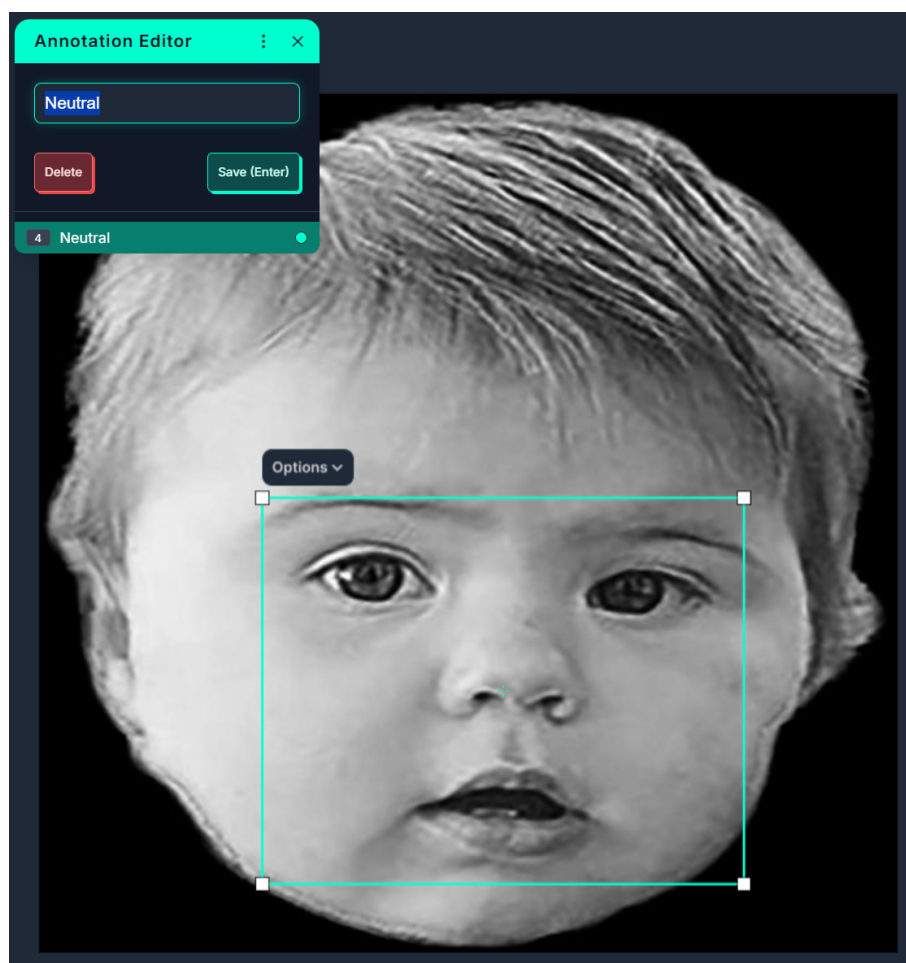


FIGURE 2.4 – Capture d'écran de l'interface d'annotation Roboflow.

Conclusion partielle

Ce chapitre a permis de détailler l'architecture globale de notre système, les fonctionnalités clés de l'application mobile, ainsi que les outils utilisés pour son développement. Le recours

à Flutter assure une expérience utilisateur moderne et fluide, tandis que Roboflow a simplifié la création d'un dataset personnalisé adapté à notre problématique.

Dans le chapitre suivant, nous aborderons les choix algorithmiques, l'entraînement du modèle IA, ainsi que l'implémentation technique détaillée de chaque composant.