

## Problématique

La crise de la COVID-19 a révélé l'importance des masques N95, dont l'efficacité dépend d'un ajustement correct pour éviter les fuites.

Toutefois :

- ✗ Les contrôles manuels et les équipements spécialisés sont coûteux et rarement accessibles.
- ✗ Les interfaces classiques sont peu intuitives pour guider correctement l'utilisateur.
- ✗ Processus lent et dépendant d'un superviseur expert humain.
- ✗ L'isolation et l'évaluation des masques par imagerie infrarouge demeurent peu exploitées. [1]

## Objectifs

Le projet vise à concevoir une station intelligente et autonome capable de segmenter les masques respiratoires sans intervention humaine. [2]

- ✓ Proposer une interaction réaliste rendant l'acquisition de vidéos infrarouges intuitive.
- ✓ Réduire la dépendance à la supervision humaine grâce à un avatar interactif.
- ✓ Entraîner un modèle de segmentation des masques à partir des images infrarouges.
- ✓ Garantir la robustesse et la généralisation du modèle à différents utilisateurs et types de masques.

## Méthodologie suivie

Notre système est organisé en deux phases :

### I- Interaction avec l'avatar et acquisition IR

Cette phase vise à guider l'utilisateur lors de la capture vidéo afin d'obtenir des images exploitables pour l'analyse thermique.

### Étape 1 – segmentation phonétique

- Le script d'instructions est segmenté afin d'assurer une synchronisation précise.
- Un audio naturel est généré à l'aide d'un modèle Text-to-Speech (TTS) pour chaque segment.
- Synchronisation labiale pour un ajustement

précis des mouvements des lèvres à chaque segment vocal en utilisant le modèle Wav2Lip.

### Étape 2 – Amélioration visuelle (Real-ESRGAN)

- Les vidéos générées à basse résolution (96×96 px) sont traitées image par image avec Real-ESRGAN, qui restaure les détails fins du visage et corrige les flous et artefacts, en particulier autour des lèvres et des contours faciaux.

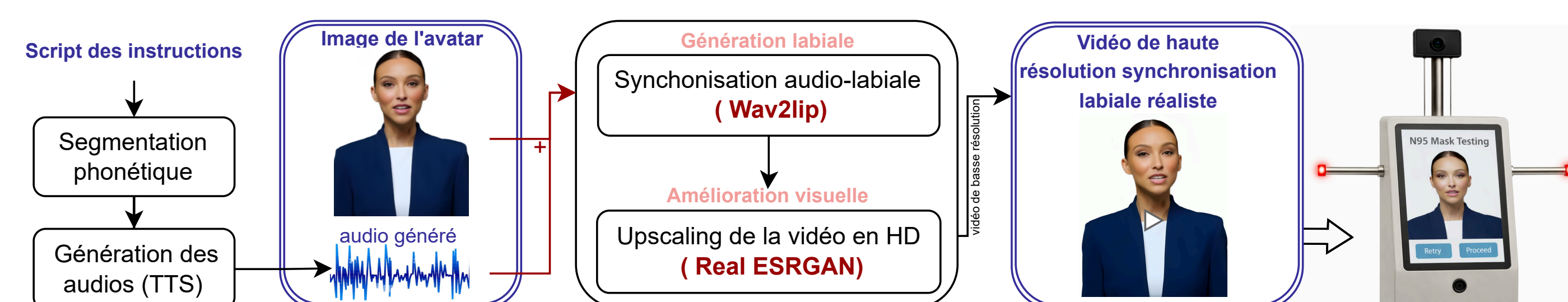


FIGURE 1 – Diagramme du système correspondant à la phase 1

## II – Segmentation des masques N95

Une fois l'acquisition effectuée, une segmentation est appliquée afin d'obtenir une isolation du masque.

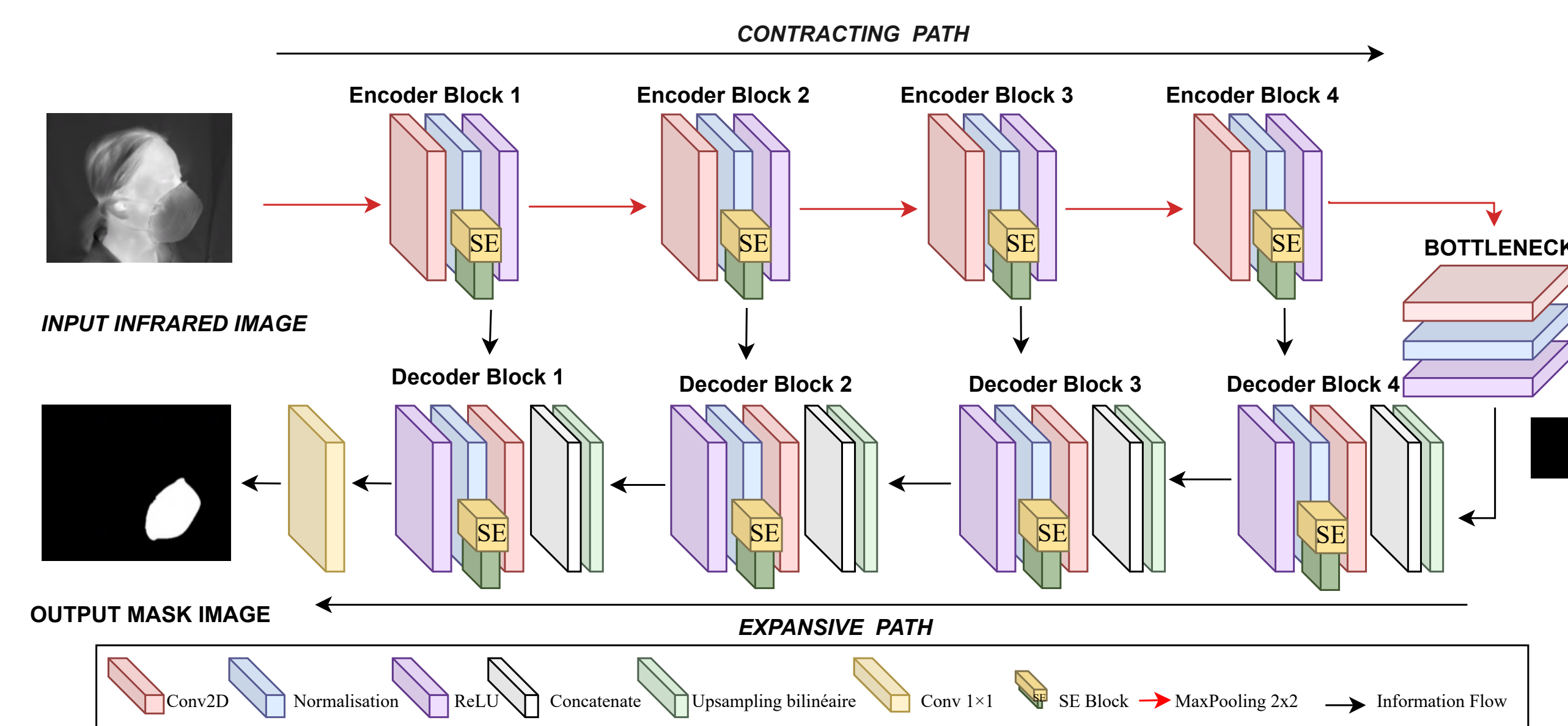


FIGURE 2 – Architecture proposée pour la segmentation des masques faciaux N95.

### Étape 3 : Segmentation avec U-Net optimisé

- Les convolutions transposées ont été remplacées par un Bilinear Upsampling afin de réduire les artefacts de type *checkerboard* et d'améliorer la qualité des contours.
- Des blocs Squeeze-and-Excitation ont été intégrés afin d'introduire un mécanisme d'attention canal et de renforcer les caractéristiques pertinentes.
- Une normalisation de Batches a été appliquée après les convolutions afin de stabiliser l'apprentissage et d'accélérer la convergence.
- Les skip connections ont été exploitées afin de préserver les informations fines.
- La fonction de perte combinée (Binary Cross-Entropy + Dice) a été utilisée afin d'optimiser la classification pixel par pixel et de mieux gérer le déséquilibre des classes.

## Résultats

Nous avons comparé notre système amélioré de la Phase 1 à la version standard de Wav2Lip, en évaluant la synchronisation labiale, la qualité visuelle, et la stabilité temporelle dans la table 1.

TABLE 1 – Résultats comparatifs de la synchronisation audio-labiale, de la qualité visuelle et de la cohérence temporelle

	Synchronisation audio - labiale			Qualité visuelle et cohérence temporelle			
Métrique	Wav2Lip	Notre système	Gain	Métrique	Wav2Lip	Notre système	Gain
LSE-D ↓	10.4	9.1	-1.3	FID ↓	46.7	41.9	-4.8
LSE-C ↑	5.3	6.1	+0.8	FVD ↓	205.9	176.4	-29.5

La figure suivante et la table 2 comparent le U-Net amélioré (phase 2) au U-Net classique sur 62 participants [3], y compris sur des masques non vus.

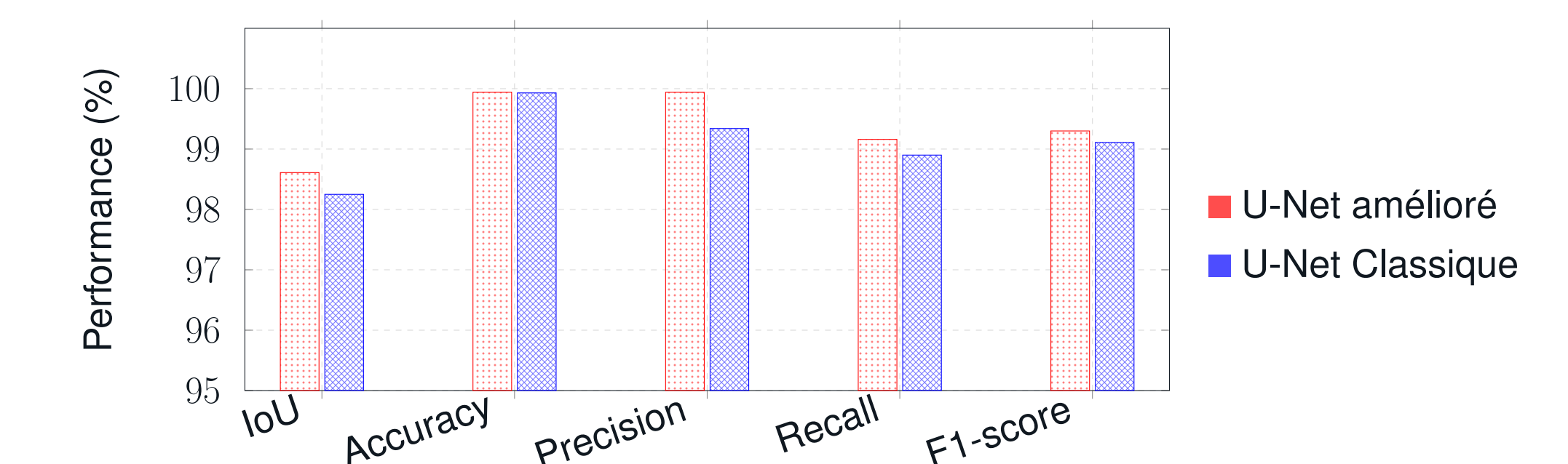


TABLE 2 – Évaluation comparative des performances des modèles

Configuration	Exactitude	Précision	Rappel	ScoreF1	IoU
U-Net standard (performance globale)	99.93	99.34	98.90	99.11	98.25
U-Net amélioré (performance globale)	<b>99.94</b>	<b>99.94</b>	<b>99.16</b>	<b>99.30</b>	<b>98.61</b>
Test sur 2200 non vu à l'entraînement	96.91	97.89	30.64	42.66	30.62
Test sur 8210 non vu à l'entraînement	97.36	98.65	48.05	59.84	47.78

## Conclusion et perspectives

✓ Wav2Lip amélioré : meilleure personnalisation d'avatar et qualité vidéo, illustrant le potentiel des avatars interactifs pour la détection de fuites.

✓ U-Net optimisé : segmentation infrarouge précise et robuste, efficace même sur des masques non vus, adaptée à la détection automatique des fuites.

## Références

- [1] Arbane et al. Advanced infrared face mask segmentation using a custom lightweight u-net model. In *2024 IEEE International Multi-Conference on Smart Systems and Green Process (IMC-SSGP)*, 2024.
- [2] Benmoussa et al. Conception d'un prototype de station préliminaire pour la vérification automatisée de l'étanchéité des masques n95, 2025.
- [3] Marchais et al. Saferespirator : comprehensive database for n95 filtering facepiece respirator leakage detection including infrared, rgb videos, and quantitative fit testing. *IEEE Access*, 2024.