

Détection de falsification vidéo

PRÉSENTATION DU STAGE DE 2ÈME ANNÉE DE CURSUS INGÉNIEUR

Encadrants : M. Christophe CHARRIER et M. Emmanuel GIGUET

Tuteur enseignant : M. Youcef IMINE

Paul CANCHON

Année universitaire 2021-2022

Plan

1. Présentation du laboratoire GREYC et de l'équipe SAFE
2. Présentation du sujet
3. Comparaison de classifieurs de détection de falsification
4. Résultats
5. Conclusion

Plan

- 1. Présentation du laboratoire GREYC et de l'équipe SAFE**
2. Présentation du sujet
3. Comparaison de classifieurs de détection de falsification
4. Résultats
5. Conclusion

Présentation du GREYC

Groupe de Recherche en Informatique, Image et Instrumentation de Caen



- Dirigé par **Christophe ROSENBERGER**
- Création en 1995 puis UMR CNRS en 2000
- Sous tutelle de l'**UNICAEN**, de l'**ENSICAEN** et du **CNRS**
- 180 membres
- Axes de recherche :
 1. Science des données
 2. Capteurs et instruments
 3. Algorithmes et IA



Equipe SAFE

Sécurité, Architecture, Forensique, biomEtrie

- Dirigée par **Christophe CHARRIER**
- Environ 30 membres dont 11 permanents
- 3 thématiques de recherche:
 1. Biométrie
 2. Architectures et modèles de sécurité
 3. Forensique



Plan

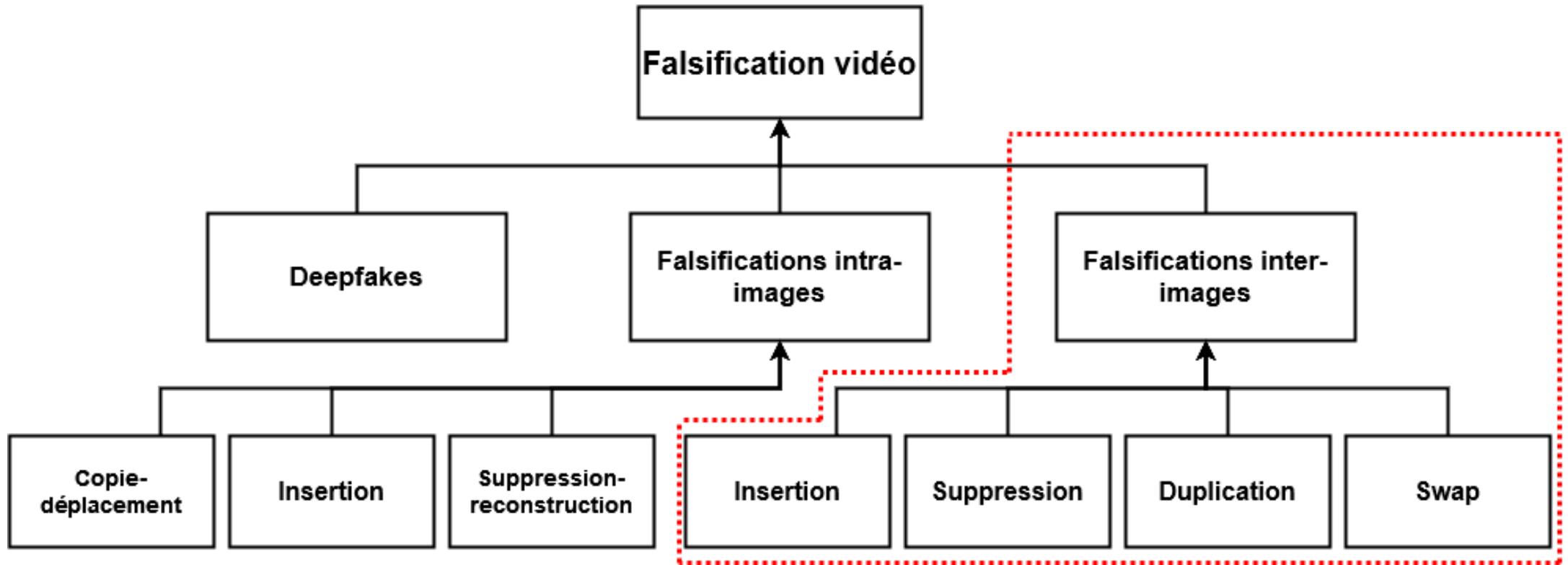
1. Présentation du laboratoire GREYC et de l'équipe SAFE
- 2. Présentation du sujet**
3. Comparaison de classifieurs de détection de falsification
4. Résultats
5. Conclusion

Contexte

- L'information est transmise beaucoup plus rapidement
- Les outils d'éditions de vidéo se démocratisent
- Multiplication des manipulations
- Besoin de pouvoir vérifier l'authenticité des vidéos



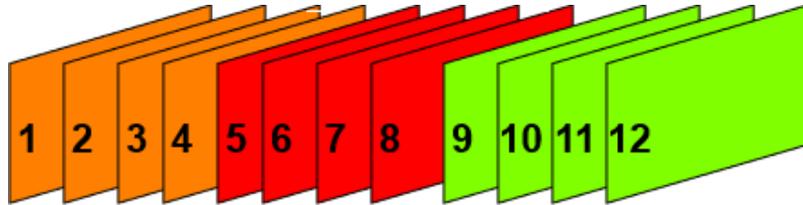
Types de falsification



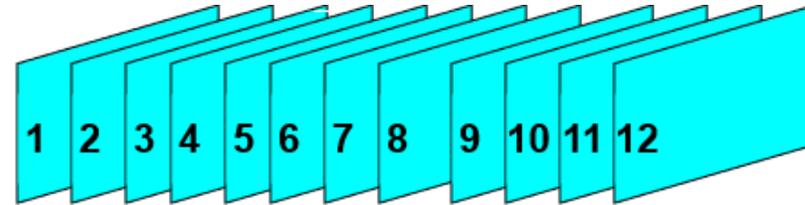
Sujet

- Comparer des méthodes de détection de falsification vidéo de type inter-images
- Particularité : Ne pas réaliser de décodage
- Vidéos utilisant le format de compression MPEG-4 AVC/H.264

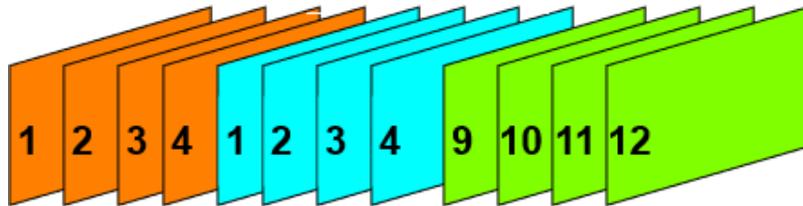
La falsification inter-images



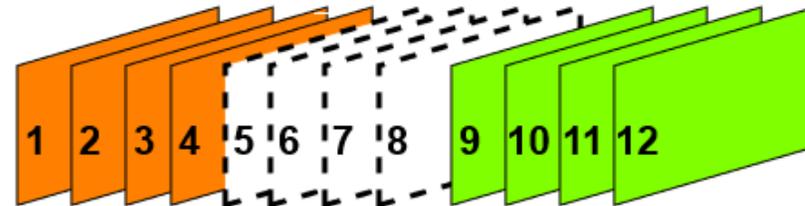
Vidéo non falsifiée A



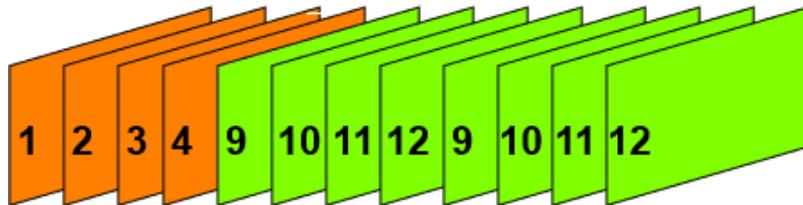
Vidéo non falsifiée B



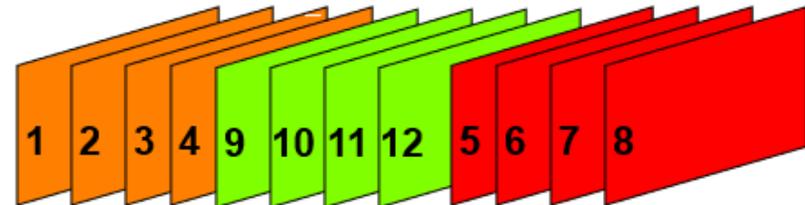
(a) Insertion



(b) Suppression



(c) Duplication



(d) Swap

Format de compression AVC/H264

- Création

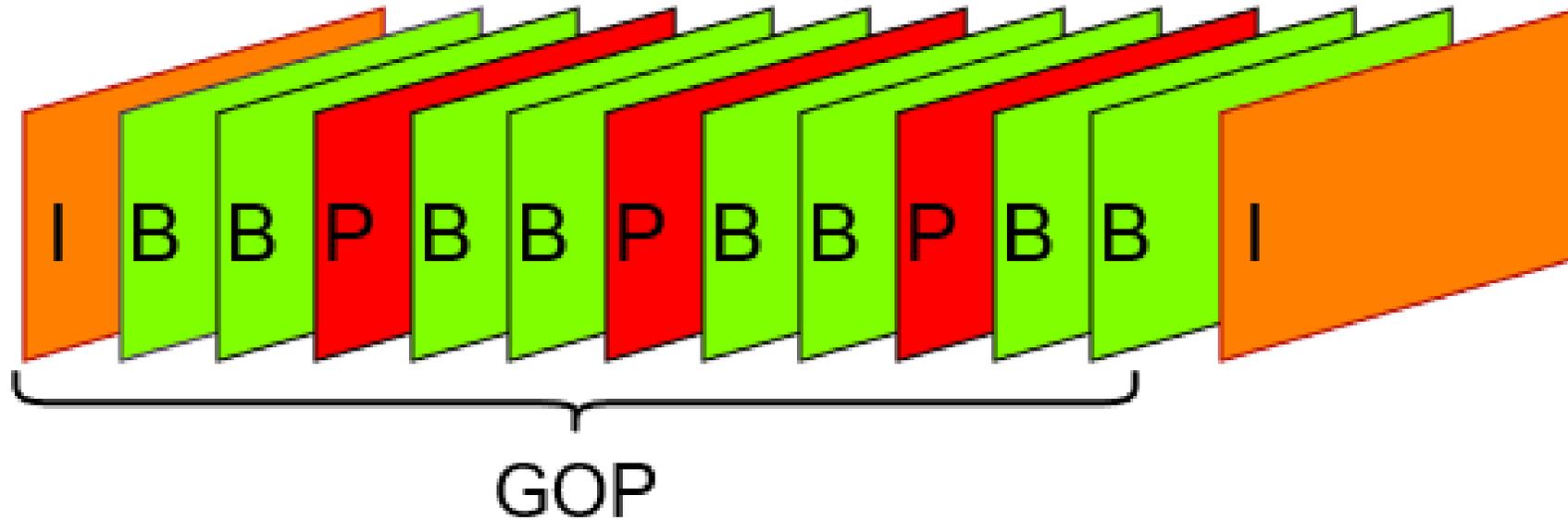
- Première version publiée en 2003



**Joint Video Team
(JVT)**

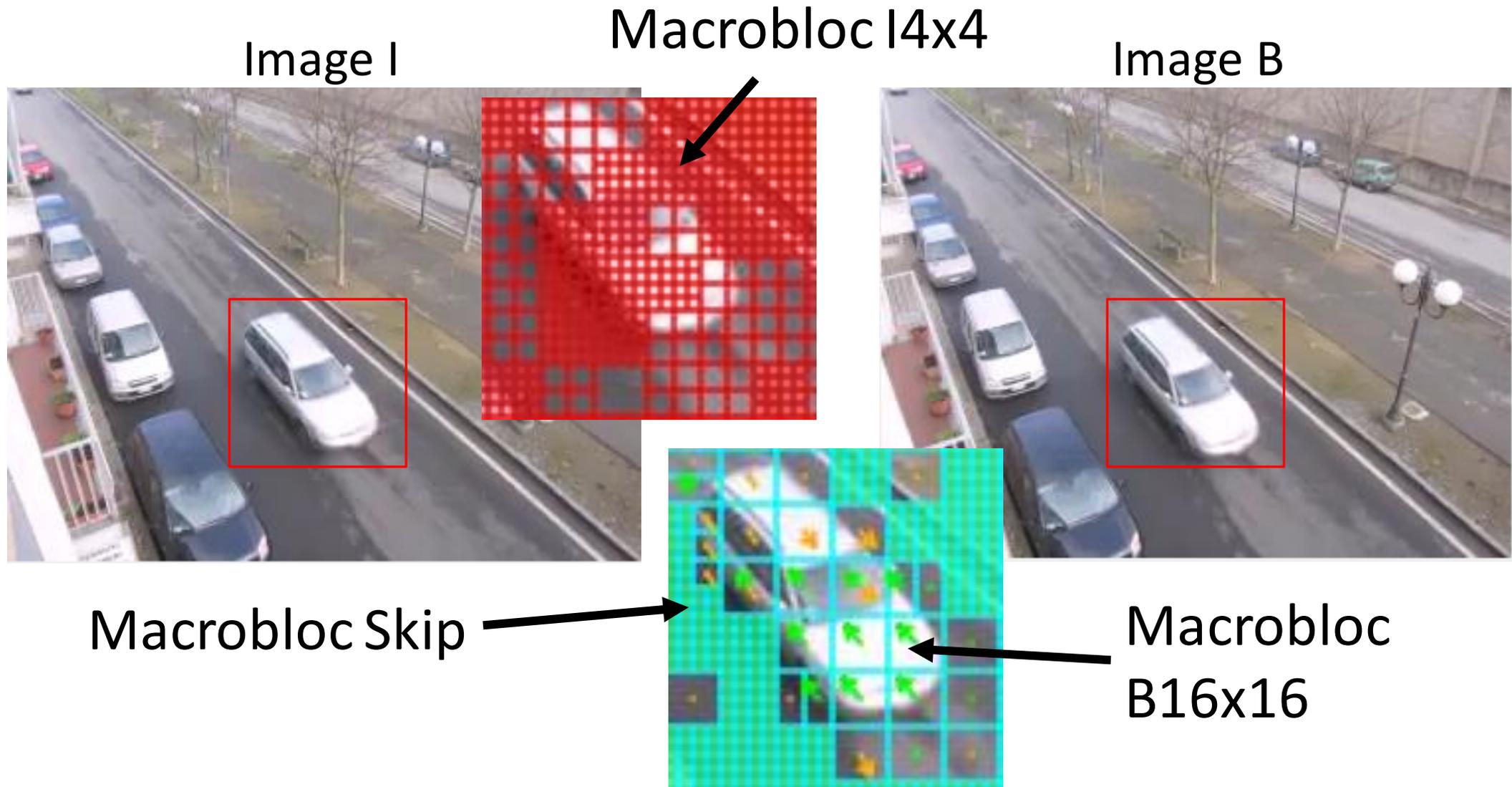
MPEG-4 AVC/H264

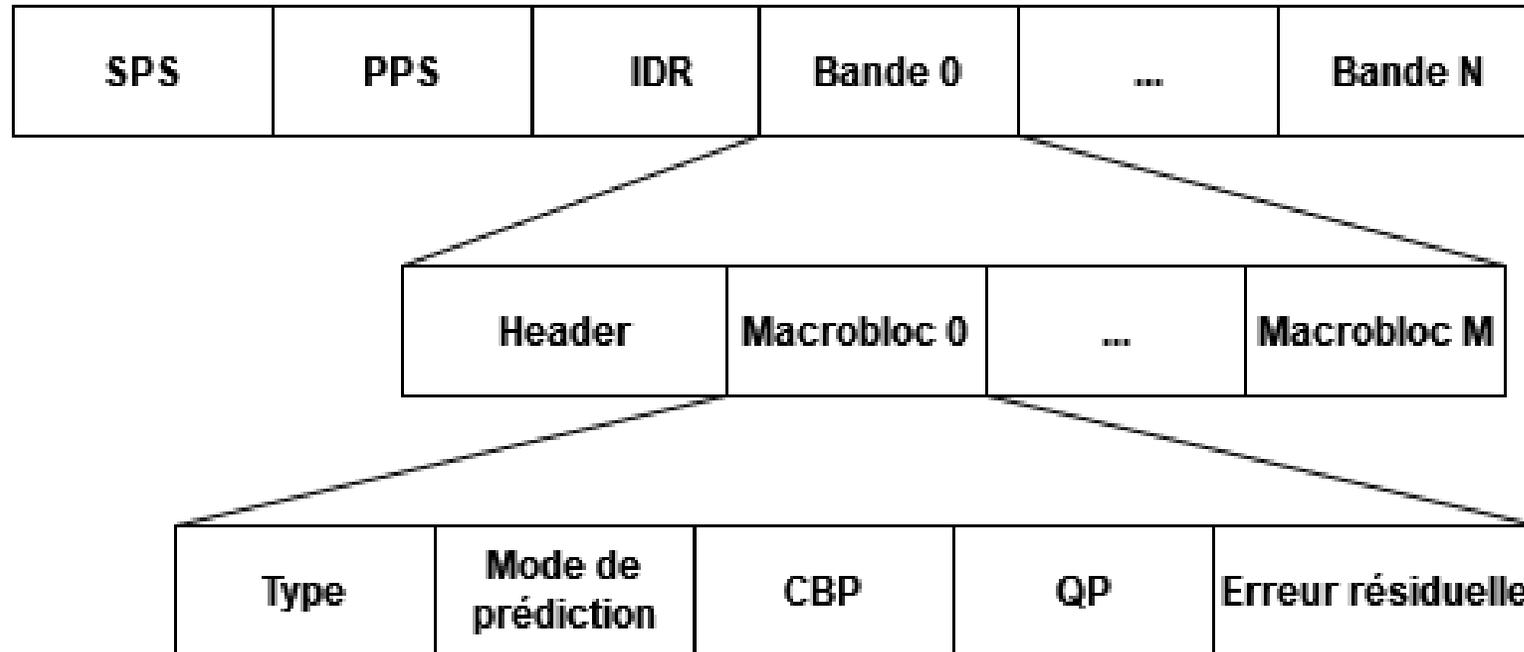
Groupes d'images (GOP)



- Image I ou image de référence
- Image P ou image prédictive
- Image B ou image bi-prédictive

Types de prédiction



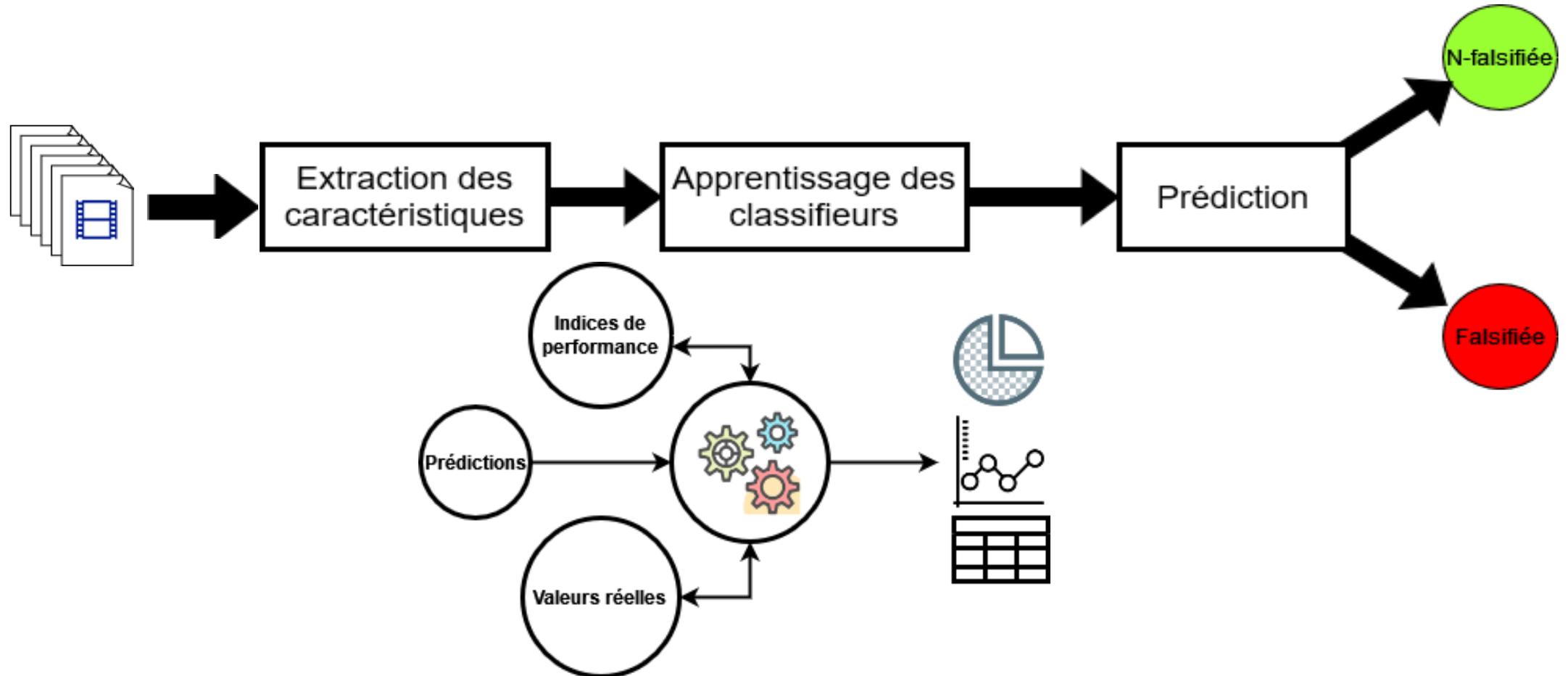


Structure du bitstream dans H.264

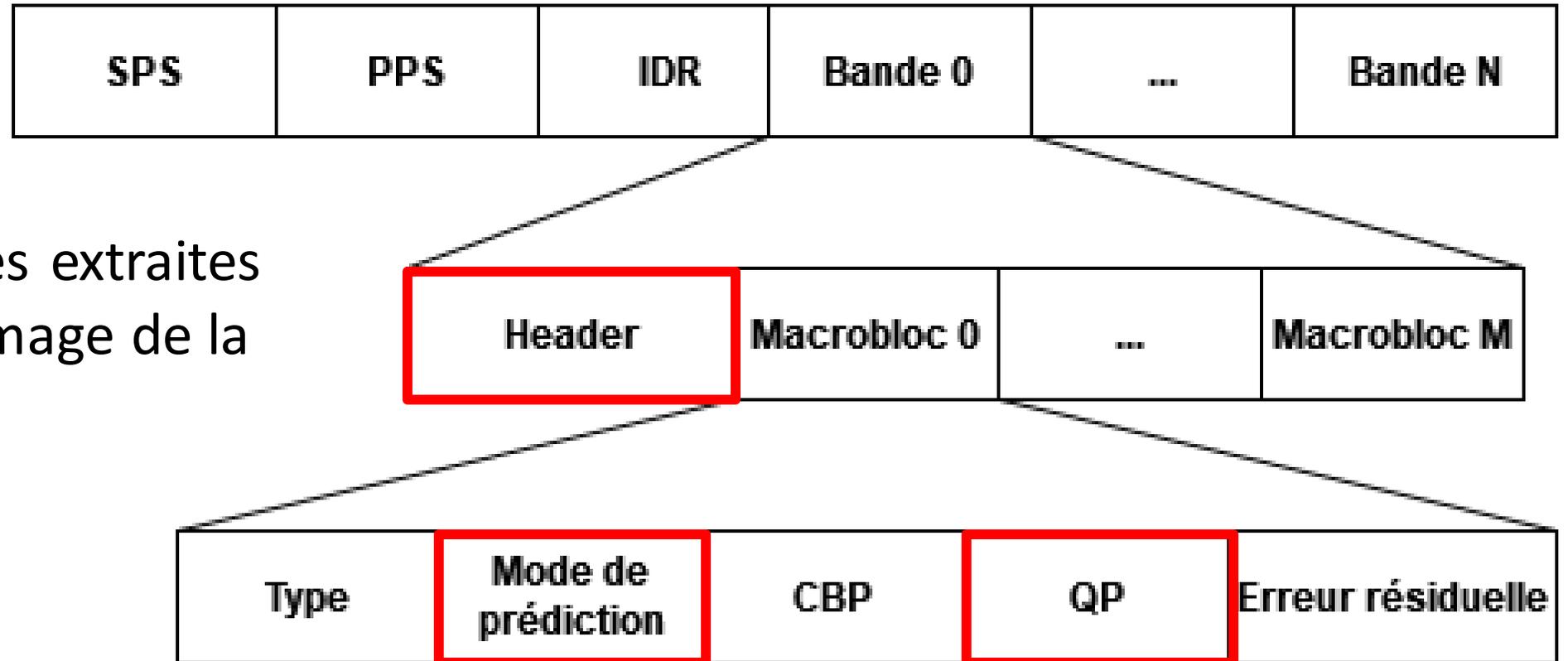
Plan

1. Présentation du laboratoire GREYC et de l'équipe SAFE
2. Présentation du sujet
- 3. Comparaison de classifieurs de détection de falsification**
4. Résultats
5. Conclusion

Schéma général du processus d'apprentissage



Extraction des caractéristiques



- Caractéristiques extraites pour chaque image de la vidéo

Extraction des caractéristiques

Extraction de 27 caractéristiques depuis le bitstream:

- Principalement basées sur les types de prédictions (20/27)

Mais aussi:

- Le *bitrate* (1/27)
- Pas de quantification *QP* et le *delta QP* (2/27)
- Longueurs des vecteurs de mouvement **moy** et **max** (4/27)

Construction des vecteurs

- Caractéristiques de chaque GOP calculées à partir des caractéristiques extraites pour chaque image
- Vecteurs construits à partir des valeurs **moy**, **min**, **max** et de l'**écart-type** de chaque caractéristique des GOPs
- Au final : 1 vidéo = 1 vecteurs de 108 dimensions ($27*4$)

Protocole d'apprentissage des classifieurs

- Création de datasets d'entraînement/de test : **80/20 %**
- Sélection aléatoire de N vidéos non-falsifiées et récupération des vidéos falsifiées associées
- Stratégie de bootstrap à **999** réplifications
- Indices de performance: **exactitude, précision, rappel, F score**

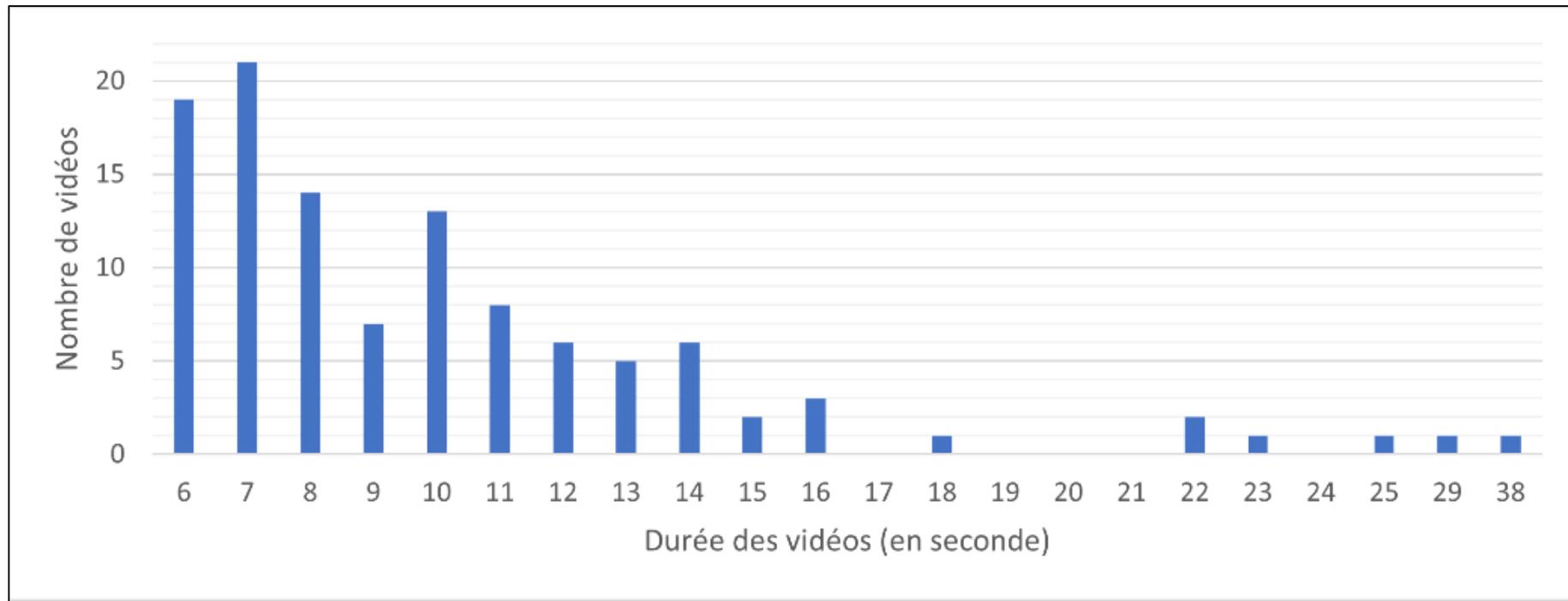
Classifieurs testés

8 classifieurs différents testés:

- SVM linéaire avec et sans pondération des classes
- SVM rbf avec et sans pondération des classes
- Régression logistique
- Classifieur bayésien naïf
- KNN
- Arbre de décision

Création de corpus

- 111 vidéos provenant de 2 corpus: **REWIND (10)** et **VIFFD (101)**

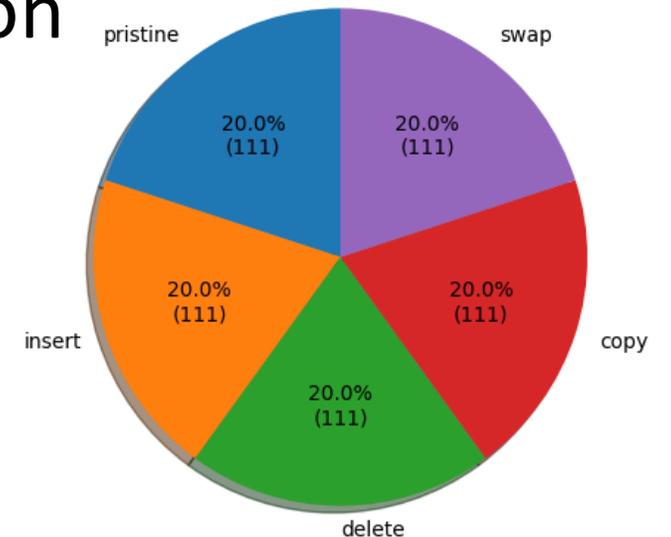


Création de corpus

- Génération automatique de falsification avec FFMPEG

1 corpus principal et 4 sous-corpus:

1. **555** vidéos : Tous les types de falsification
2. **222** vidéos : **insertion**
3. **222** vidéos : **suppression**
4. **222** vidéos : **duplication**
5. **222** vidéos : **swap**



Plan

1. Présentation du laboratoire GREYC et de l'équipe SAFE
2. Présentation du sujet
3. Comparaison de classifieurs de détection de falsification
4. **Résultats**
5. Conclusion

Evaluations sur le corpus principal

Classifieur	Exactitude	Précision	Rappel	F-score
Arbre de décision	0.78	0.89	0.82	0.85
k plus proches voisins	0.79	0.81	0.96	0.88
Régression logistique	0.89	0.93	0.94	0.93
Naives Bayes	0.62	0.84	0.65	0.73
SVM linéaire	0.90	0.94	0.93	0.94
SVM linéaire avec pondération	0.88	0.96	0.89	0.92
SVM rbf	0.89	0.93	0.93	0.93
SVM rbf avec pondération	0.86	0.95	0.88	0.91

Résultats satisfaisants pour SVM et régression logistique

Evaluations sur les sous-corpus

- Performances globalement moins bonnes que pour le corpus principal:
 - **Insertion** : 91% d'exactitude avec SVM linéaire
 - **Suppression** : 84% d'exactitude avec SVM linéaire
 - **Duplication** : 83% d'exactitude avec SVM linéaire
 - **Swap** : 86% d'exactitude avec SVM linéaire
- **Hypothèse** : Les sous-corpus contiennent uniquement 222 vidéos
Soit 44 pour les tests

Plan

1. Présentation du laboratoire GREYC et de l'équipe SAFE
2. Présentation du sujet
3. Comparaison de classifieurs de détection de falsification
4. Résultats
5. **Conclusion**

Conclusion

Personnel:

- Découverte du métier de chercheur et de la vie en laboratoire
- Méthodologie de recherche
- Développement des compétences en Python, Shell
- Perspective : écriture d'un article

Merci

Des
questions?