

# Détection et reconnaissance automatique des tirs au basketball

Rapport de projet

**Réalisé par :**

Ghanam Khalil

Sukach Danylo

**Domaine :** Intelligence artificielle, données et sport

**Année universitaire :** 2025–2026

# Table des matières

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Architecture générale du système</b>	<b>2</b>
<b>3</b>	<b>Détection des tirs par YOLOv8</b>	<b>2</b>
3.1	Rôle du modèle YOLO . . . . .	2
3.2	Détection frame-level du type de tir . . . . .	3
<b>4</b>	<b>Construction du dataset vidéo</b>	<b>3</b>
<b>5</b>	<b>Reconnaissance temporelle par CNN 3D</b>	<b>3</b>
5.1	C3D : Convolutional 3D Network . . . . .	3
5.2	I3D : Inflated 3D ConvNet . . . . .	3
<b>6</b>	<b>Résultats expérimentaux</b>	<b>4</b>
6.1	Performance globale . . . . .	4
6.2	Matrice de confusion . . . . .	4
6.3	Analyse des résultats . . . . .	4
<b>7</b>	<b>Problèmes rencontrés</b>	<b>4</b>
<b>8</b>	<b>Méthode heuristique Layup / Dunk</b>	<b>4</b>
<b>9</b>	<b>Suivi des événements de tir</b>	<b>5</b>
<b>10</b>	<b>Estimation de la distance du tir</b>	<b>5</b>
<b>11</b>	<b>Conclusion</b>	<b>5</b>

# 1 Introduction

L'analyse automatique de vidéos sportives est un domaine central de la vision par ordinateur. Elle permet d'extraire des informations à forte valeur ajoutée, telles que la reconnaissance d'actions, l'analyse de performances ou la génération de statistiques avancées.

Dans le basketball, la reconnaissance automatique des types de tirs constitue un défi majeur en raison de la rapidité des actions, de la similarité visuelle entre certains gestes (notamment entre layup et dunk), de la variabilité des angles de caméra et du déséquilibre naturel des données disponibles.

L'objectif de ce projet est de concevoir un système capable de détecter et classifier automatiquement les principaux types de tirs :

- Jump shot,
  - Layup,
  - Dunk.
- 

## 2 Architecture générale du système

Le système proposé repose sur une approche hybride combinant apprentissage profond et règles heuristiques. L'idée principale est de séparer clairement :

- l'analyse spatiale (où se trouvent les objets ?),
- l'analyse temporelle (comment le mouvement évolue ?),
- et l'analyse géométrique (distance, hauteur relative).

Le pipeline global est organisé comme suit :

1. Détection spatiale des objets et actions (YOLOv8),
  2. Reconnaissance temporelle des actions (CNN 3D),
  3. Heuristiques pour distinguer Layup et Dunk,
  4. Suivi des événements de tir,
  5. Rendu et visualisation.
- 

## 3 Détection des tirs par YOLOv8

### 3.1 Rôle du modèle YOLO

YOLOv8 est utilisé comme module de perception visuelle. Il effectue une détection image par image des objets et actions pertinentes du basketball, notamment :

- le joueur,
- le ballon,
- le panier,
- le joueur effectuant un jump shot,
- le joueur effectuant un layup ou un dunk,
- le ballon dans le panier.

YOLO fournit ainsi des indices spatiaux précis, mais ne permet pas à lui seul une reconnaissance fiable du type de tir.

### 3.2 Détection frame-level du type de tir

À chaque frame, le type de tir candidat est déterminé selon les classes détectées :

- *player-jump-shot* → Jump shot,
- *player-layup-dunk* → Layup ou Dunk,
- *ball-in-basket* → Tir réussi.

Afin de réduire le bruit et les erreurs ponctuelles, une fenêtre temporelle glissante est utilisée pour déterminer l'action dominante sur plusieurs frames consécutives.

—

## 4 Construction du dataset vidéo

Les vidéos sont découpées en clips courts de 4 à 6 secondes, centrés sur les actions de tir détectées.

La répartition du dataset est la suivante :

Classe	Train + Validation	Test	Total
Layup	1290	240	1530
Jump shot	1290	240	1530
Dunk	90	20	110

Cette distribution met en évidence un déséquilibre important, particulièrement pour la classe Dunk.

—

## 5 Reconnaissance temporelle par CNN 3D

### 5.1 C3D : Convolutional 3D Network

C3D est un réseau convolutionnel utilisant des convolutions 3D, capables de capturer simultanément l'apparence spatiale et la dynamique temporelle des actions. Il traite une vidéo comme un volume spatio-temporel et apprend des motifs liés aux mouvements du joueur et du ballon.

Cependant, C3D reste sensible au déséquilibre des classes et montre des limites pour distinguer des actions très proches visuellement, comme le layup et le dunk.

### 5.2 I3D : Inflated 3D ConvNet

I3D étend des réseaux 2D vers la vidéo en inflatant les filtres dans la dimension temporelle. Cette approche permet un apprentissage plus stable et une meilleure capture des mouvements complexes.

Malgré ces avantages, I3D reste limité lorsque les actions diffèrent principalement par des critères géométriques, tels que la distance au panier ou la hauteur du saut.

—

## 6 Résultats expérimentaux

### 6.1 Performance globale

Le modèle I3D atteint une accuracy de :

$$\text{Accuracy} = 48.8\%$$

### 6.2 Matrice de confusion

	Layup	Dunk	Jump Shot
Layup	28	15	197
Dunk	0	19	1
Jump shot	29	19	202

### 6.3 Analyse des résultats

Les résultats montrent une forte confusion entre les classes Layup et Jump Shot. Le modèle tend à prédire Jump Shot dans les situations ambiguës, ce qui indique qu'il se base principalement sur la posture de tir, sans intégrer explicitement la distance ou la hauteur.

—

## 7 Problèmes rencontrés

Plusieurs limites ont été identifiées :

- prédiction excessive de la classe Jump Shot,
  - confusion entre Layup et Dunk,
  - insensibilité à la distance réelle du tir,
  - déséquilibre important des classes,
  - absence d'informations géométriques explicites.
- 

## 8 Méthode heuristique Layup / Dunk

Pour pallier les limites des CNN 3D, une méthode heuristique basée sur la géométrie a été introduite.

Lorsque YOLO détecte une action de type *player-layup-dunk*, la position du joueur est comparée à celle du panier :

- **Dunk** : la tête ou les mains du joueur atteignent ou dépassent la hauteur du panier,
- **Layup** : le joueur est plus bas et utilise généralement le panneau.

Cette approche permet de corriger partiellement les erreurs de classification.

—

## 9 Suivi des événements de tir

Chaque tir est modélisé comme un événement structuré comportant trois états :

- START : début du tir,
- MADE : tir réussi (ballon détecté dans le panier),
- MISSED : tir raté (absence de réussite après un délai).

Chaque événement est enregistré avec sa position, sa distance, son type et son résultat.

—

## 10 Estimation de la distance du tir

La distance du tir est estimée à l'aide d'une transformation de perspective permettant de convertir les coordonnées image en coordonnées terrain réelles. La distance euclidienne entre la position du joueur et le panier est ensuite calculée.

Cette estimation reste sensible aux angles de caméra et à la qualité de la détection des éléments du terrain.

—

## 11 Conclusion

Ce projet met en évidence les limites des approches purement apprenantes pour la reconnaissance fine des actions sportives. Les résultats obtenus montrent qu'un modèle CNN 3D seul ne suffit pas à distinguer des actions proches comme le layup et le dunk.

L'approche hybride combinant détection par YOLO, reconnaissance temporelle par CNN 3D et heuristiques géométriques permet d'améliorer la robustesse et l'interprétabilité du système. Ce travail ouvre la voie à des améliorations futures, notamment par l'intégration de données multi-vues et de contraintes physiques plus explicites.