# Keys of Imagination

## Sarim Aeyzaz and Ehtsham Walidad
### FAST NUCES

## Precise Information about the Proposal

The "Keys of Imagination" project leverages computer vision and deep learning to transform a simple paper drawing of a piano keyboard into an interactive digital piano. The system uses hand-tracking, fingertip detection, and YOLO-based object detection to identify key presses and play corresponding musical notes.
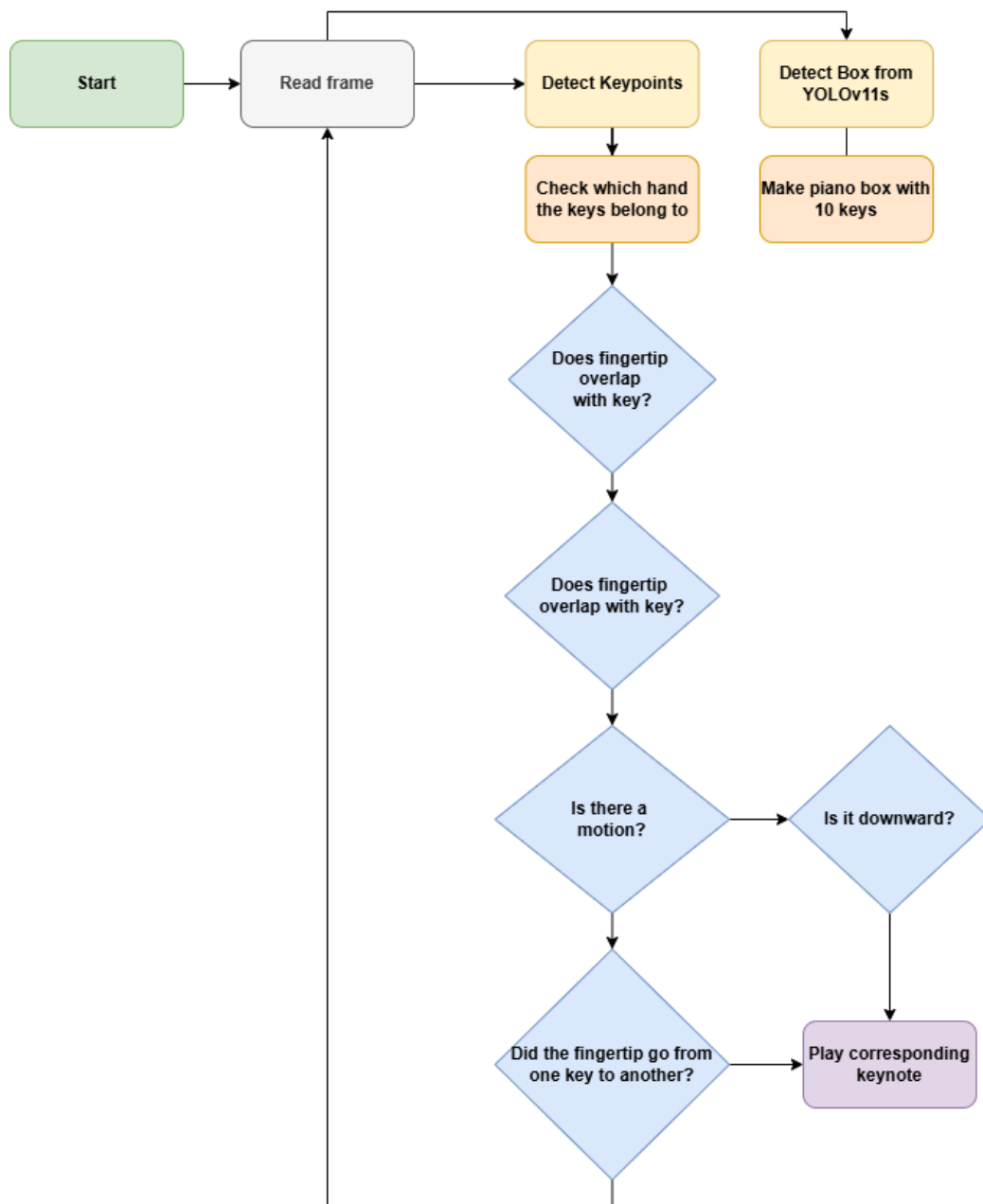
### Contributions to Your Work

- Developed a system that integrates MediaPipe and YOLO for robust hand and object detection.

- Implemented motion-based keypress detection with noise reduction through buffer zones and cooldowns.

- Enhanced functionality to support two-hand interactions for a richer user experience.

- Transitioned from screen-based keyboards to real-world object detection using YOLO.

## Methodology

The methodology involves six incremental stages of development, starting from basic fingertip detection to integrating YOLO for real-time piano detection. Below is a flow diagram and the corresponding explanation of each stage.

# Flow Diagram

```
┌─────────┐      ┌────────────┐      ┌──────────────────┐      ┌────────────────────┐
│  Start  │ ───▶ │ Read frame │ ───▶ │ Detect Keypoints │      │ Detect Box from    │
└─────────┘      └────────────┘      └──────────────────┘      │ YOLOv11s           │
                                              │                 └────────────────────┘
                                              ▼                            │
                                     ┌──────────────────┐      ┌────────────────────┐
                                     │ Check which hand │      │ Make piano box with│
                                     │ the keys belong to│      │ 10 keys            │
                                     └──────────────────┘      └────────────────────┘
```

**Start** → **Read frame** → **Detect Keypoints** → **Check which hand the keys belong to**

**Detect Box from YOLOv11s** → **Make piano box with 10 keys**

Decision nodes (diamonds):

- **Does fingertip overlap with key?**
- **Does fingertip overlap with key?**
- **Is there a motion?** → **Is it downward?**
- **Did the fingertip go from one key to another?** → **Play corresponding keynote**

**Is it downward?** → **Play corresponding keynote**

# Implementation Details

## Model Selection

- **MediaPipe**: Used for hand keypoint detection due to its efficiency and accuracy.

- **YOLO (You Only Look Once)**: Used for detecting the paper piano keyboard, chosen for its real-time performance.

## Incremental Stages of Development

- **Stage 1: Basic Keypress Detection**

  - MediaPipe detected hand keypoints.
  - A keypress was registered if a fingertip (index, middle, ring, or pinky) was below other keypoints.

- **Stage 2: On-Screen Keyboard**

  - Displayed a virtual keyboard on the screen.
  - Played a note when a fingertip overlapped with a key.
  - Introduced a buffer/deadzone ( 10%) to reduce misdetections on key boundaries.
  - Prevented repeated notes by tracking which finger was on which key.

- **Stage 3: Motion-Based Key Detection**

  - Detected downward motion (y-coordinate increase) for keypress registration.
  - Implemented a 0.5-second cooldown to avoid spamming notes.
  - Addressed transition issues between keys (fixed in Stage 4).

- **Stage 4: Two-Hand Interaction**

  - Extended functionality to detect keypresses for both hands.
  - Fixed a bug where transitions between keys were not played correctly under the new motion logic.

- **Stage 5: Contour-Based Bounding Box Detection**

– Used contours to create a bounding box around the detected keyboard for more precise key segmentation.

- **Stage 6: YOLO Integration**

  – Integrated YOLO to detect the paper piano keyboard in real-time.
  – Ensured the system ignored other objects (e.g., the webcam).

## Hyperparameters

- **MediaPipe Minimum Detection Confidence**: 0.7

- **MediaPipe Minimum Tracking Confidence**: 0.7

- **YOLO Confidence Threshold**: 0.2

- **Cooldown Time**: 0.5 seconds

- **Buffer Margin**: 10% of key dimensions

- **History Length**: 5 frames

## Results and Analysis

- The system successfully identified key presses and played notes in real-time.

- The YOLO integration significantly improved detection robustness compared to earlier contour-based methods.

- Two-hand interaction allowed simultaneous chord playing, enhancing musical expression.

- Limitations were observed in cases of image occlusion, particularly when hands overlapped the keys.

# Strengths of Your Work

- Real-time performance with minimal latency.

- Integration of motion-based detection reduces false positives.

- Support for two-hand interactions enhances usability and realism.

- Modular development allows future extensions (e.g., additional instruments).

# Limitations of Your Work

The primary limitation of the system is image occlusion, where overlapping hands or poor lighting conditions can affect detection accuracy.

# Conclusion

The "Keys of Imagination" project successfully transforms a simple drawing into an interactive musical instrument. Future work could address occlusion issues by incorporating depth sensors or advanced segmentation techniques.