

# A Estrutura da Proposta

A estrutura da **Proposta do Projeto** entregue (observe o prazo no Plano de Ensino da disciplina) é:

## 1. Título do projeto

Desenvolvimento de um Sistema de Reconhecimento de Gestos Dinâmicos em Tempo Real Usando Temporal Transformers e Esqueletos 2D

## 2. Nomes dos participantes da equipe

Davi Baechtold Campos

## 3. Nome e assinatura do orientador

Orientador: Prof. Dr. Alceu de Souza Brito Junior

Co-orientador: Prof. Dr. Alessandro Zimmer

## 4. Motivação para o desenvolvimento do projeto

A crescente integração de tecnologia no cotidiano demanda interfaces humano-máquina (IHM) mais seguras e intuitivas. Em contextos como o automotivo, onde a atenção do motorista é crítica, ou no controle de robôs, métodos tradicionais de interação (telas, botões) podem ser inadequados. O reconhecimento de gestos surge como uma solução promissora, porém, a interpretação de movimentos dinâmicos em tempo real a partir de câmeras comuns (monoculares) ainda é um desafio técnico. Este projeto propõe o desenvolvimento de um sistema robusto e de baixo custo para essa finalidade, utilizando a arquitetura *Temporal Transformer*, estado da arte em modelagem de sequências, para classificar gestos a partir de sequências de esqueletos 2D, com potencial de aplicação direta nas indústrias automotiva e de robótica.

## 5. Objetivo

Desenvolver e avaliar um sistema de reconhecimento de gestos dinâmicos em tempo real, baseado em sequências de esqueletos 2D extraídas de vídeo monocular, utilizando um modelo *Temporal Transformer*.

Para alcançar este objetivo principal, os seguintes objetivos específicos foram definidos:

- Revisar o estado da arte em reconhecimento de gestos e selecionar uma base de dados pública para o projeto.
- Implementar um pipeline para extração e normalização de esqueletos 2D em tempo real com MediaPipe.
- Construir e treinar um modelo *Transformer* para classificar as sequências de gestos.
- Avaliar o desempenho do modelo com métricas de acurácia e F1-score.
- Desenvolver uma aplicação de demonstração em tempo real.
- Analisar a aplicação do sistema em casos de uso para o controle de interfaces automotivas e robóticas.

## 6. Descrição sucinta das teorias e técnicas utilizadas

1. **Seleção e Pré-processamento de Dados:** Será realizada uma revisão de bases de dados públicas, como a **HAGRID (HAnd Gesture Recognition Image Dataset)**, que é adequada por não ser específica de uma língua de sinais. O conjunto de dados escolhido será formatado e dividido em subconjuntos de treinamento, validação e teste.
2. **Extração de Características:** Será utilizado o framework **MediaPipe**, com seu modelo **BlazePose** (pré-treinado na base COCO), para extrair em tempo real os 33 pontos-chave (landmarks) do corpo e mãos. A

saída será uma sequência temporal de coordenadas 2D para cada gesto.

3. **Modelagem e Treinamento:** As sequências de landmarks normalizadas alimentarão um modelo **Temporal Transformer**, implementado em **PyTorch**. A arquitetura usará mecanismos de **auto-atenção** para identificar as poses mais relevantes em cada gesto e **codificação posicional** para entender a ordem dos movimentos.
4. **Avaliação e Aplicação:** O modelo treinado será avaliado quantitativamente e integrado a uma aplicação de demonstração em tempo real usando **OpenCV** para validar sua eficácia e aplicabilidade nos casos de uso propostos.

## 7. Descrição do hardware e/ou do software necessários para o desenvolvimento

- **Hardware:** Computador com no mínimo 8GB de RAM, GPU NVIDIA com suporte a CUDA (para treinamento) e webcam. Os recursos são próprios do aluno.
- **Software:** SO Linux, Python 3.8+, PyTorch, OpenCV, MediaPipe, NumPy e Scikit-learn.

## 8. Contexto do Projeto

Este TCC, de natureza acadêmica, insere-se nas áreas de IHM e Visão Computacional. A iniciativa originou-se durante um intercâmbio na THI (Technische Hochschule Ingolstadt) em um projeto focado em automação veicular. O projeto explorará dois casos de uso principais: controle de interfaces na indústria automotiva e controle de braços robóticos, com potencial de integração com o laboratório de robótica da PUCPR.

## 9. Referências bibliográficas básicas

ALMJALLY, Abrar; ALMUKADI, Wafa Sulaiman. Deep computer vision with artificial intelligence based sign language recognition to assist hearing and speech-impaired individuals. **Scientific Reports**, v. 15, n. 1, p. 32268, 2025.

JIANG, Songyao et al. Skeleton Aware Multi-modal Sign Language Recognition. **arXiv preprint arXiv:2103.08833**, 2021.

LI, Dongxu et al. Transferring Cross-domain Knowledge for Video Sign Language Recognition. In: **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition**. 2020. p. 6204-6213.

SILVA, Karolayne Teixeira da. **SignWriting para Reconhecimento de Gestos em Língua de Sinais**. Dissertação (Mestrado em Ciência da Computação) – Centro de Informática, Universidade Federal de Pernambuco, Recife, 2025.

VASWANI, Ashish et al. Attention is all you need. In: **Advances in neural information processing systems**. 2017. p. 5998-6008.

Assinaturas

Prof. Dr. Alceu de Souza Brito Junior

Orientador

**Importante: A proposta deve ter, no máximo, duas páginas e ser assinada pelo orientador.**