

Nota Conceitual e Plano de Implementação do Projeto Capstone

Título do Projeto: Detecção de Mudanças e Degradação de Habitats por Imagens de Satélites para o Ecoturismo Angolano

Membros da Equipe:

Israel Cassute

Jose Ndala

Nota Conceitual

1. Visão Geral do Projeto

O projeto **INZO-HABITAT-ML** visa combater a rápida degradação de habitats (desflorestação, erosão, expansão não planeada) em áreas de conservação vitais, como o Parque Nacional da Quiçama, que são cruciais para o Ecoturismo Angolano.

A solução é o desenvolvimento de um sistema automatizado de **Segmentação Semântica** usando Aprendizado Profundo (Deep Learning - DL) para classificar o uso e a cobertura do solo a partir de imagens de satélite Sentinel-2. O resultado é uma **ferramenta de alerta de detenção precoce** que fornece dados geoespaciais precisos aos gestores de conservação.

Relevância para os ODS: O projeto contribui diretamente para o **ODS 15 (Vida na Terra)**, ao monitorizar ecossistemas e identificar a perda de conservação (Meta 15.1), e para o **ODS 13 (Ação Climática)**, ao detetar a degradação de sumidouros de carbono.

2. Objetivos

1. **Desenvolvimento:** Construir e treinar um modelo **U-Net** de Segmentação Semântica multi-classe (5 classes: Floresta, Degradação, etc.) usando 9 bandas espectrais do Sentinel-2.
2. **Desempenho:** Alcançar um **Mean IoU (Média de Interseção sobre União)** superior a **0.70** no conjunto de validação, com foco na precisão da classe minoritária (Degradação/Solo Exposto).
3. **Entrega:** Implementar um *pipeline* automatizado de inferência para processar novas imagens de satélite e gerar mapas de degradação acionáveis (acionáveis) em tempo quase real.

3. Histórico

A monitorização da degradação em Angola é tradicionalmente baseada em levantamentos de campo ou classificadores supervisionados básicos (e.g., Random Forest), que são lentos, caros e limitados na precisão *pixel-a-pixel*. A literatura demonstra que a complexidade espectral e

espacial das imagens de satélite modernas (Sentinel-2, 9 bandas) requer uma técnica mais avançada. O **Aprendizado de Máquina (DL)** se torna benéfico porque a arquitetura **U-Net** pode extrair padrões hierárquicos complexos, superando as limitações dos métodos tradicionais na tarefa de identificar e mapear com precisão os limites irregulares da degradação.

4. Metodologia

A metodologia centra-se no **Deep Learning** para a tarefa de **Segmentação Semântica**.

- **Modelo Principal: U-Net** - Selecionada pela sua capacidade de combinar contexto global (Encoder) com detalhes de alta resolução (Skip Connections), essencial para segmentar limites de classes irregulares como a erosão/desflorestação.
- **Treinamento:** Supervisionado, utilizando a função de perda **Combined Loss (Categorical Cross-Entropy + Dice Loss)** para mitigar o desbalanceamento inerente entre a classe majoritária (Floresta) e a classe minoritária (Degradação).
- **Saída:** Um mapa de probabilidade de 5 canais (One-Hot Encoded), onde cada pixel é classificado em uma das cinco classes de interesse.

5. Diagrama de Projeto de Arquitetura

Componente	Função	Tecnologia Principal
1. Fonte de Dados	Coleta e <i>stacking</i> de imagens Sentinel-2 (GeoTIFF, 9 bandas).	Google Earth Engine API / Sentinelsat
2. Pré-Processamento	Normalização de reflectância (0-1.0), divisão em <i>patches</i> de 128x128 pixels, e Codificação OHE.	Python / Rasterio / NumPy
3. Modelo (U-Net)	Executa o treinamento e a inferência. O núcleo do Aprendizado Profundo.	TensorFlow / Keras
4. Serviço (API)	Recebe requisições de novas coordenadas/imagens e envia para o modelo.	Flask / FastAPI

5. Pós-Processamento	Agrega os <i>patches</i> inferidos de volta em um GeoTIFF completo e aplica regras de alerta.	Python / GDAL
6. Interface de Usuário	Visualização interativa dos mapas de degradação e alertas para gestores de conservação.	Web App (Streamlit ou GIS Web)

6. Fontes de Dados

- **Fonte:** Satélite **Sentinel-2** da Agência Espacial Europeia (ESA).
- **Tipo:** Imagens **Multiespectrais** (L2A - Corrigido Atmosfericamente), resolução de 10m, 9 bandas espectrais.
- **Dados de Treinamento:** Conjuntos de imagens recortadas para a área de estudo (e.g., Parque Nacional da Quíçama) e correspondentes **máscaras de Segmentação (*Ground Truth*)** rotuladas manualmente para as 5 classes.
- **Pré-processamento:** Normalização (divisão por 10000.0) e *Patching* (janelamento) para treino eficiente.

7. Literatura de Apoio

O projeto está solidamente ancorado nas pesquisas que demonstram a eficácia do Deep Learning para a classificação ambiental. Trabalhos como os de **Ma et al. (2029)** e **Zhu et al. (2017)** estabelecem que as Redes Neurais Convolucionais (CNNs) são superiores aos classificadores tradicionais na extração de padrões espaciais em dados de Sensoriamento Remoto. Nosso trabalho baseia-se diretamente nesta fundação, alavancando a arquitetura U-Net para **aumentar a precisão na localização dos limites de degradação** (Segmentação Semântica), que é um avanço além da simples classificação de píxeis.

Plano de Implementação

1. Pilha de Tecnologia

Componente	Tecnologia	Função
Linguagem de Programação	Python	Linguagem principal para ML, SR e <i>backend</i> de API.

Frameworks ML/DL	TensorFlow / Keras	Definição, treinamento e serialização da U-Net.
Bibliotecas Geoespaciais	Rasterio / GDAL	Leitura, escrita e manipulação de GeoTIFFs.
Computação Científica	NumPy / Scikit-learn	Manipulação de <i>arrays</i> (patches) e cálculo de métricas (IoU).
Estrutura de Serviço	Flask / FastAPI	Criação do <i>endpoint</i> RESTful para servir as previsões do modelo.
Hardware	Cloud Compute (GPU/TPU)	Aceleração essencial para o treinamento da U-Net.

2. Linha do Tempo (Modelo de 8 Semanas)

Tarefa Principal	Fases	Duração Estimada
Fase 1: Preparação de Dados	Coleta de Imagens, Rotulagem e Geração de Patches 128x128.	2 Semanas (S1-S2)
Fase 2: Desenvolvimento de Modelo Base	Definição da Arquitetura U-Net, Treinamento Inicial (50 épocas).	2 Semanas (S3-S4)
Fase 3: Refinamento e Otimização	Ajuste de Hiperparâmetros (LR, otimizador), Otimização da Loss Function (Dice Loss) para Degradação.	2 Semanas (S5-S6)

Fase 4: Implantação e Teste Final	Serialização, Desenvolvimento da API (Flask/FastAPI) e Teste de Inferência de ponta a ponta em GeoTIFFs não vistos.	2 Semanas (S7-S8)
--	---	-------------------

(Se tiver um par, adicione uma Matriz de Distribuição de Tarefas aqui.)

3. Marcos

- **M1 (Semana 2):** Conclusão do Conjunto de Dados de Treinamento e Validação rotulados (100% dos *patches* rotulados e prontos para o *pipeline* de treino).
- **M2 (Semana 4):** Modelo Base Treinado e Avaliado com **Mean IoU > 0.65**.
- **M3 (Semana 6):** Modelo Otimizado com **Mean IoU > 0.70** e *Recall* melhorado para a classe 'Degradação'.
- **M4 (Semana 8):** *Pipeline* de Inferência implementado, testado e acessível via **API funcional**, gerando mapas GeoTIFF segmentados e acionáveis.

4. Desafios e Mitigações

Desafio Potencial	Estratégia de Mitigação
Qualidade dos Dados (Nuvens)	Usar apenas produtos Sentinel-2 L2A (corrigidos) e compósitos de imagens (mediana mensal) para filtrar a cobertura de nuvens.
Desempenho do Modelo (Bias de Classe)	Implementar a Dice Loss para dar maior peso à classe minoritária 'Degradação' e aplicar <i>data augmentation</i> para simular variações de vegetação.
Restrições Técnicas (Tempo de Treino)	Utilizar a capacidade de GPU/TPU de plataformas <i>Cloud</i> e, se necessário, explorar <i>backbones</i> de rede mais leves para a U-Net.

5. Considerações Éticas

- **Preconceito (Bias):** O modelo pode apresentar *bias* se for aplicado a um ecossistema (e.g., floresta de miombo) significativamente diferente daquele em que foi treinado (e.g., Quiçama).

- **Ação:** Documentação clara das limitações do modelo e a exigência de **calibração/retreinamento** ao ser aplicado a novas regiões ecológicas.
- **Impacto na Comunidade Alvo:** Os dados sobre desflorestação/erosão podem levar a intervenções governamentais que afetem o uso tradicional da terra pelas comunidades locais.
 - **Ação:** Priorizar a **transparência dos dados** e garantir que o projeto seja apresentado como uma **ferramenta de apoio e não de fiscalização coerciva**, promovendo a consulta e a capacitação dos *stakeholders* locais no uso da tecnologia para a gestão sustentável.

6. Referências

Artigos Científicos e Livros:

- **Belgiu, M., & Drăguț, L. (2016).** *Random forest in remote sensing: A review of applications and future directions*. ISPRS Journal of Photogrammetry and Remote Sensing, 114, 24-31.
 - **Canty, M. J. (2019).** *Image analysis, classification and change detection in remote sensing* (3rd ed.). CRC Press.
 - **Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., et al. (2013).** *High-resolution global maps of 21st-century forest cover change*. Science, 342(6160), 850-853.
 - **Lu, D., Mausel, P., Brondízio, E., & Moran, E. (2004).** *Change detection techniques*. International Journal of Remote Sensing, 25(12).
 - **Ma, L. et al. (2029).** *Deep Learning in Remote Sensing Applications: A Meta-Analysis and Review*. ISPRS Journal of Photogrammetry and Remote Sensing, 152, 166-177.
 - **Zhu, X. X. et al. (2017).** *Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources*. IEEE Geoscience and Remote Sensing Magazine, 5(4), 8-36.
-