

KINTSUGI

Shredded Document Reconstruction via Pairwise Seam Classification

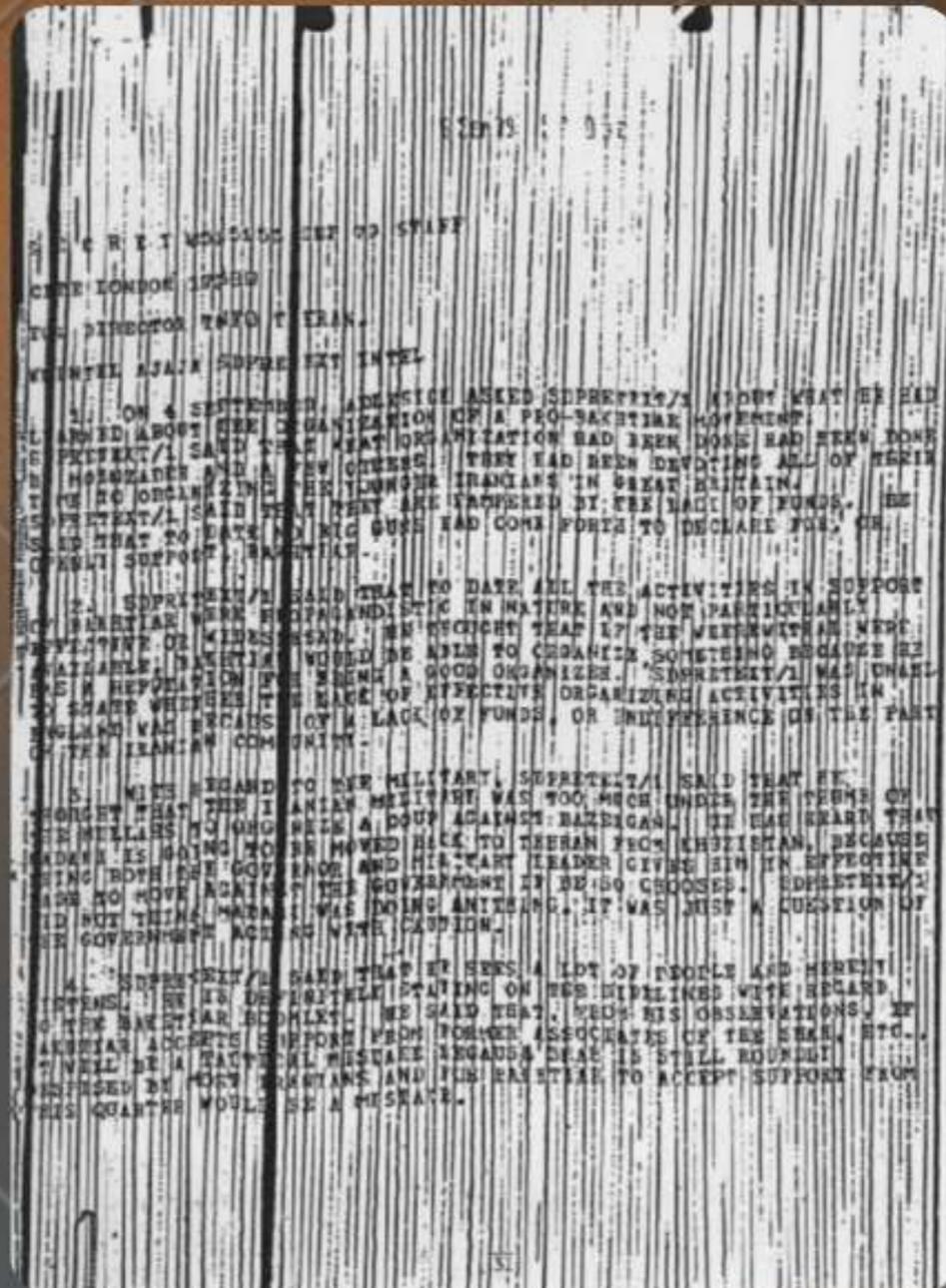
Anggota Kelompok 14:

Cathlyn S. D.

Clarissa A.

Dara O.

Ryan W.



PROBLEM & OBJECTIVES

The challenge of **Shredded Document Reconstruction** is critical in forensics and archival science, but manual assembly is impossible for high volumes.

DEV & TECH STACK

- PYTORCH,
- STREAMLIT,
- OPENCV,
- NUMPY,
- PILLOW,
- PDF2IMAGE/POPPLER,
- PLOTLY,
- MATPLOTLIB,
- SEABORN,
- HUGGINGFACE

Libraries and framework

- 2XRTX 3070, AMD EPYC 7B12 64-CORE
PROCESSOR, 128 GB OF DDR4 RAM,
- RENTED FROM VAST.AI
 - USED TERMINAL SSH AND FILEZILLA

Hardware configuration



KINTSUGI
LIVE DEMO

[HTTP://CODENAME-
KINTSUGI.STREAMLIT.APP/](http://codename-kintsugi.streamlit.app/)

COMPARISON WITH OTHER APPROACHES

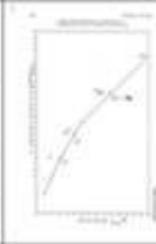
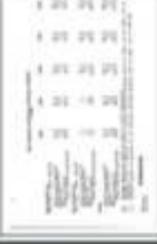
Approach Type	Pairwise Matching Method	Global Assembly Method	Key Advantages	Disadvantages & Limitations
1. Traditional Computer Vision	Raw pixels / Simple edge features.	Optimal Search (MCHPP/TSP) or Greedy.	Simple to implement, fast <i>pairwise</i> evaluation.	Pairwise Matching Fails: Extremely sensitive to noise and alignment; cannot handle strip-cut noise.
2. Feature-Based Methods	Handcrafted features (contours, text lines).	Often use MCHPP to ensure global optimum.	More robust than pixel matching, especially for strip-cut.	Feature Extraction Fails: Cannot generalize to different fonts or layouts; less robust against cross-cut noise.
Pairwise Seam Classification	Deep CNN Classifier for robust "Match/No Match" scoring.	Multi-start Greedy Heuristic	Highest Accuracy & Robustness for pairwise scoring, successfully handles complex seam features (text/line continuity).	Greedy Assembly Risk: Not guaranteed to find the absolute optimal sequence, potentially leading to cumulative errors in long documents. Primarily limited to vertical strip-cut documents.

DATASET USED

Kami menggunakan dataset RVL-CDIP yang tersedia di [HuggingFace](https://huggingface.co/datasets/robertoribon/rvl-cdip). Dataset ini terdiri dari 400.000 gambar yang terbagi dalam 16 kategori (seperti surat, formulir, invoice), dengan distribusi seimbang 25.000 gambar per kelas. Ukurannya, ~35GB.

400_k

Total Images
(320K TRAIN, 40K VAL&TEST)

Questionnaire	Letter	Form	Email	Handwritten	Scientific Report	ScientificPublication	Specification
							
Advertisement	Budget	File Folder	News Article	Invoice	Presentation	Resume	Memo
							

16 Categories
(classes)

25_k Images /class

DATA PREPROCESSING

Streaming

Memuat dataset RVL-CDIP sebesar 38GB secara real-time via HuggingFace. Jadi datanya ditarik langsung dari cloud per-batch, tanpa perlu download manual.

Dynamic Pair Generation

Melatih AI pakai 3 skenario :

- Positive Sample - 50% Peluang (Code: if choice < 0.5)
- Hard Negative - 25% Peluang (Code: elif choice < 0.75)
- Easy Negative (Code : Else)

Image Stitching

```
combined = np.concatenate((strip1, strip2), axis=1)
```

Proses penggabungan (concatenation) dua potongan kertas terpilih (Kiri & Kanan) menjadi satu gambar utuh agar bisa dilihat oleh model.

ResNet Augmentation

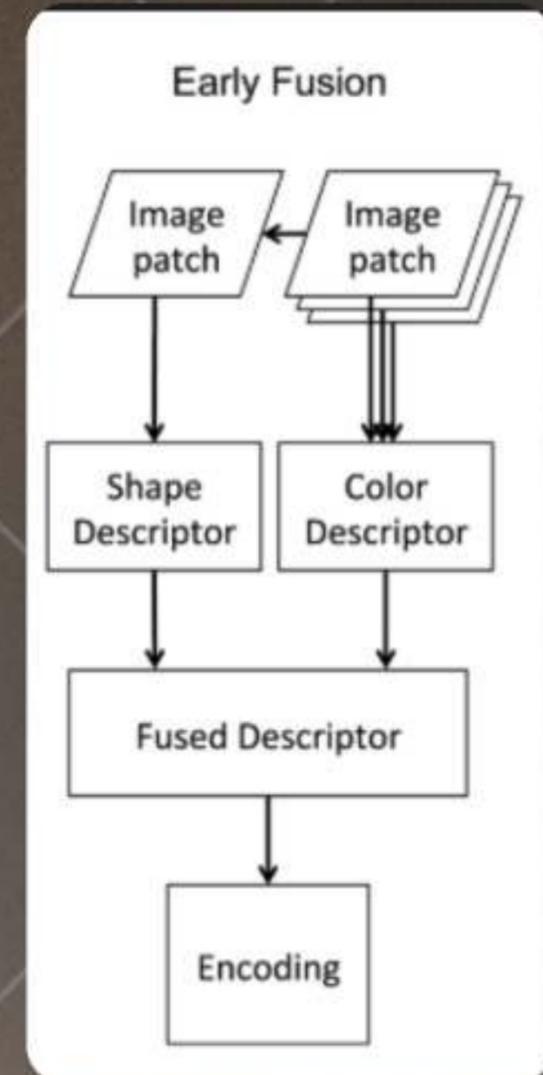
- Random Crop 224x224
- Grayscale (pola, garis sambungan, tesktur)
- Normalize

PIXEL-LEVEL CONTINUITY W/ EARLY FUSION

Early Fusion → Data input digabungkan pada tahap paling awal pipeline yang merepresentasikan “kandidat sambungan” secara utuh.

Joint Feature Learning → Receptive field convolution filter pertama langsung mencakup kedua strip secara bersamaan.

- **Deteksi Micro-Texture:** Menganalisis spatial coherence pada tingkat piksel.
- **Seam Boundary Analysis**
 - Match → Gradien intensita piksel halus
 - No Match → Diskontinuitas Mendadak



ADVERSARIAL TRAINING W/ MINING

Translation Invariance



Limitasi random negatives → tidak efektif karena yang dipelajari adalah perbedaan global texture variance, BUKAN presisi posisi.

HARD NEGATIVE MINING → membuat data yang secara semantical semiliar (berasal dari area sama) tetapi secara spatial inaccurate.

- Memaksa Loss Function untuk mempenalti kesalahan alignment sekecil apapun, sehingga model mempelajari **strict pixel alignment &** robust terhadap **translational invariance**

SEAMRESNET: RESNET-18 + CUSTOM DENSE

Backbone: ResNet-18

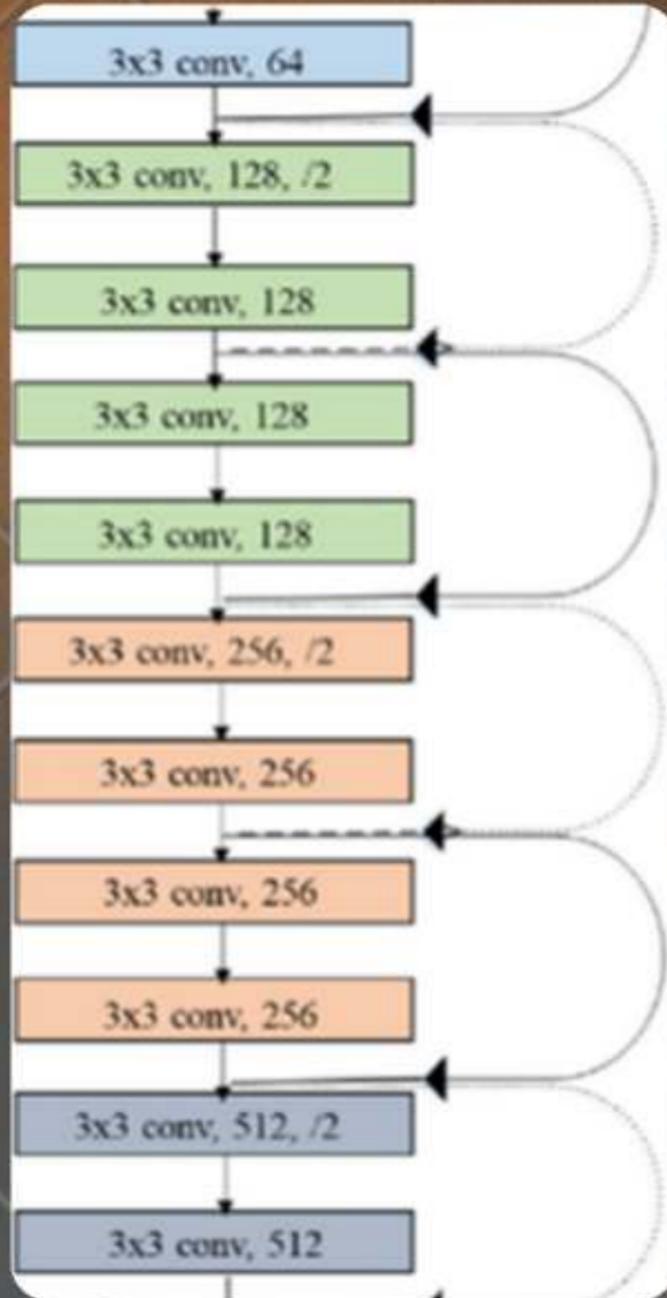
- Diinisialisasi dengan ImageNet Weights.
- Memberikan akses ke low-level feature extractors (seperti Gabor Filters) yang efektif untuk edge & texture detection.

Custom Classification Head (Dense Block)

- Arsitektur klasifikasi standar diganti dengan blok fully connected untuk output biner (includes feature compression, regularization, binary logit, etc.)

Inference

- Output Logit diproses melalui Sigmoid Activation untuk menghasilkan probabilitas biner (0, 1)



TRAINING LOOP

Honorable Mentions:

- Melakukan on-the-fly augmentation (Random Cropping & Greyscale)
 - Optimizer: AdamW with LR=1e-4
 - Loss: BCSWithLogitsLoss (Sigmoid & Binary Cross Entropy)
 - Batch Size = 256; Epochs = 10; Steps per epoch: 3000

ACCURACY: 82.7%

SORRY LUPA SS, GK ADA FOTO BUKTI FOR NOW :(

LIMITATIONS

Shred Type Restriction

The system is currently limited to vertical strip-cut documents.

Assembly Algorithm Bottleneck

The global reassembly relies on a Multi-start Greedy Heuristic.

Data Generalization

The model may perform poorly on documents containing color, complex textures, or photographs

THANK YOU

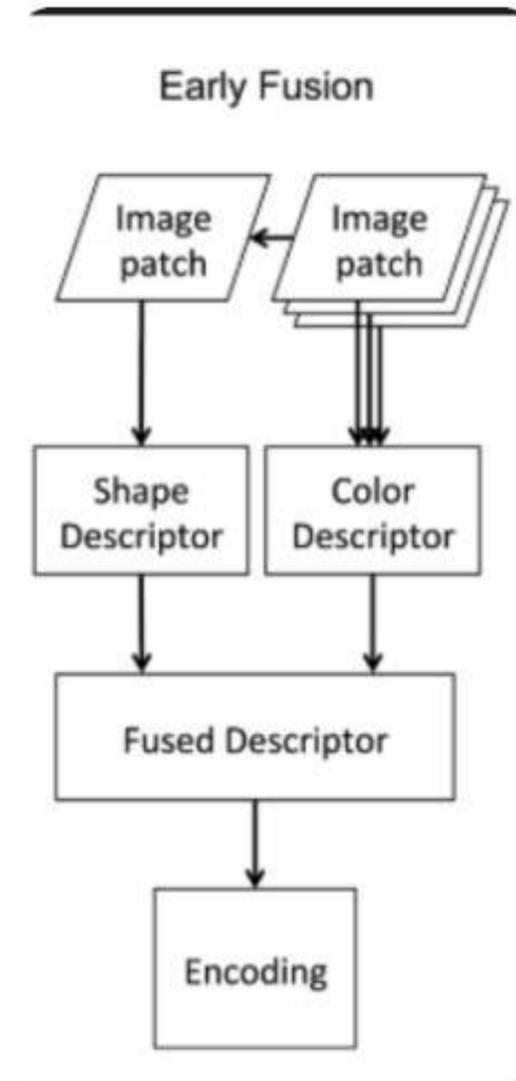
KELOMPOK 14 - LG01

PIXEL-LEVEL CONTINUITY W/ EARLY FUSION

Early Fusion → Data input digabungkan pada tahap paling awal pipeline yang merepresentasikan “kandidat sambungan” secara utuh.

Joint Feature Learning → Receptive field convolution filter pertama langsung mencakup kedua strip secara bersamaan.

- **Deteksi Micro-Texture:** Menganalisis spatial coherence pada tingkat piksel.
- **Seam Boundary Analysis**
 - Match → Gradien intensita piksel halus
 - No Match → Diskontinuitas Mendadak



ADVERSARIAL TRAINING W/ MINING

Translation Invariance



Limitasi random negatives → tidak efektif karena yang dipelajari adalah perbedaan global texture variance, BUKAN presisi posisi.

HARD NEGATIVE MINING → membuat data yang secara semantical semiliar (berasal dari area sama) tetapi secara spatial inaccurate.

- Memaksa Loss Function untuk mempenalti kesalahan alignment sekecil apapun, sehingga model mempelajari **strict pixel alignment &** robust terhadap **translational invariance**