Project Log: Removing the ”Green Goo”

Goo Removal

Gradient Descendants

December 4, 2025

Overview

The dataset contains synthetic green artifacts (”goo”) overlaying the tissue samples. These must be removed to prevent the neural network from learning them.

Phase 1: The Naive Approach

Attempt 1: Simple Thresholding

Hypothesis: Since the artifacts are green and the tissue is pink/purple, a simple HSV color check should work.

We defined a standard green range in HSV space:

- Lower: $[35, 100, 50]$
- Upper: $[85, 255, 255]$

The Result (Failure): While it caught the center of the goo, it failed at the edges. The synthetic goo has **anti-aliasing** (semi-transparent edges). The simple threshold missed these faint pixels, leaving a distinct ”green halo” or ring around the removal area. This would result in dirty labels.

Phase 2: The Brute Force Fix

Attempt 2: Aggressive Dilation

Hypothesis: If the mask is slightly too small (missing the halo), let's just force it to be bigger using Morphological Dilation.

We applied ‘cv2.dilate‘ with ‘iterations=5‘.

```
# The Brute Force Logic
kernel = np.ones((3, 3), np.uint8)
# Expand blindly in all directions
dilated_mask = cv2.dilate(goo_mask, kernel, iterations
    =5)
```

The Result (Failure): This solved the halo problem but introduced a worse one: **Data Destruction**. Because dilation expands blindly in every direction, it expanded the mask *into* the valid tissue. On images where the goo was touching tissue cells, we ended up deleting actual biological data.

Phase 3: The Final Solution

Final Solution: Core & Shell Logic

Epiphany: We need a way to detect the faint halo (like Attempt 2) but stop expanding exactly where the green tint ends (unlike Attempt 2).

We implemented a **”Smart Connection”** algorithm inspired by Hysteresis Thresholding.

Step A: Define Two Masks

Instead of one threshold, we use two:

1. **The Core (Strict):** High Saturation (> 100). Captures the solid center.
2. **The Shell (Loose):** Low Saturation (> 30). Captures the halo, but also captures noise in the tissue.

Step B: Connected Components

We cannot use the Shell mask directly because it is noisy. **The Logic:** We only keep a blob from the ”Shell” mask if it is physically touching a blob from the ”Core” mask.

- Isolated faint green blob → Tissue Noise (Discard).
- Faint green blob touching Core → Halo (Keep).

Step C: The 1-Pixel Expansion

Even with the smart logic, a sub-pixel fringe often remained. We added a single iteration of dilation at the very end. Since the mask is now ”clean” (no noise), expanding it by 1 pixel is safe and ensures the anti-aliasing is fully covered.

Conclusion

The **Core & Shell + Expansion** method successfully removes 100% of the artifact (including the halo) while preserving the integrity of significant amount of adjacent tissue structures.

Appendix A: Mathematical Formulation of the "Core & Shell" Algorithm

Let $I : \Omega \to \mathbb{R}^3$ be the input image defined on the discrete domain $\Omega \subset \mathbb{Z}^2$. We first transform the image from the RGB color space to the HSV color space, denoted as $I_{hsv}(x) = (H(x), S(x), V(x))$ for any pixel $x \in \Omega$.

The artifact removal pipeline is modeled as a **Geodesic Reconstruction** (a morphological transformation that reconstructs objects from a "marker" image based on their connectivity within a "mask" image).

1. Hysteresis Thresholding

We define the artifact segmentation based on two distinct predicates, forming a subset relationship $M_{core} \subset M_{shell}$.

Let $\mathcal{H}_{green} = [35, 85]$ be the range of Hue values corresponding to the green artifact. Let $\tau_{high} = 100$ and $\tau_{low} = 30$ be the saturation thresholds.

We define the **Core Marker** (M_{core}) as the set of pixels satisfying the strict condition:

$$M_{core} = \{x \in \Omega \mid H(x) \in \mathcal{H}_{green} \land S(x) > \tau_{high}\} \quad (1)$$

Plain English: *Select pixels that are definitely solid green. This identifies the undeniable center of the artifact, ignoring edges and noise.*

We define the **Shell Mask** (M_{shell}) as the set of pixels satisfying the relaxed condition:

$$M_{shell} = \{x \in \Omega \mid H(x) \in \mathcal{H}_{green} \land S(x) > \tau_{low}\} \quad (2)$$

Plain English: *Select pixels that are even faintly green. This captures the semi-transparent "halo" around the artifact, but also mistakenly captures yellowish spots in the tissue.*

2. Topological Filtering (Reconstruction)

The set M_{shell} contains both the valid artifact halo and disjoint background noise (false positives). To isolate the artifact, we apply a connectivity constraint using **Connected Components** (graph-based algorithm that identifies distinct, isolated "blobs" of pixels).

Let $\mathcal{C}(S)$ denote the set of connected components of a binary set S. For any component $C_i \in \mathcal{C}(M_{shell})$, we retain it if and only if it intersects with the Core marker:

$$M_{rec} = \bigcup_{C_i \in \mathcal{C}(M_{shell})} \{C_i \mid C_i \cap M_{core} \neq \emptyset\} \quad (3)$$

Plain English: *Look at every "faint green blob." If it is physically touching a "solid green blob," keep it. If it is floating alone in the tissue, delete it.*

This operation effectively reconstructs the full shape of the artifact from the high-confidence markers, filtering out all C_j where $C_j \cap M_{core} = \emptyset$.

3. Boundary Regularization (The "Nudge")

Due to interpolation artifacts in the source image generation, a sub-pixel fringe often exists outside the color detection range of M_{shell}.

Let B be a flat, unitary structuring element of size 3×3. We define the final artifact mask M_{final} as the **Morphological Dilation** (an operation that expands the boundaries of foreground pixels) of the reconstructed set:

$$M_{final} = M_{rec} \oplus B = \{z \in \Omega \mid (B_z \cap M_{rec}) \neq \emptyset\} \quad (4)$$

Plain English: *Expand the final cleaned mask outward by exactly 1 pixel in all directions. This covers the invisible, microscopic color fringe at the very edge to ensure the training data is perfectly clean.*