



L'utilizzo dell'intelligenza artificiale nel settore agroalimentare

Prof. Fabrizio Sarghini

Ing. Angela De Vivo, Ph.D.

Ing. Mariano Crimaldi, Ph.D.



Il ruolo del Machine Learning (ML) nell'Industria Alimentare

Uno dei fondamenti dell'Intelligenza Artificiale è il Machine Learning (ML), ovvero l'apprendimento automatico. L'idea centrale è che un sistema possa essere addestrato a svolgere un compito, imparando dai dati, senza che debba essere programmato in modo esplicito per ogni possibile situazione. Durante la fase di addestramento, si forniscono all'algoritmo dati in ingresso (input) e il risultato atteso (output). L'algoritmo, attraverso un processo iterativo, ottimizza i propri parametri fino a riuscire a riconoscere schemi nascosti e regole all'interno dei dati.

Nel Food, il Machine Learning (ML) sfrutta i tanti dati raccolti (dalla produzione alla distribuzione) per:

**Prevedere con
maggiore
accuratezza rischi
e trend di mercato**



**Ottimizzare
processi produttivi
riducendo sprechi
di risorse**



**Migliorare la
sicurezza e la
qualità dei prodotti
alimentari**



**Gestire filiere
complesse,
monitorando in tempo
reale ogni fase “dal
campo alla tavola”**



Le menti dell'IA: panoramica sulle tipologie di ML

Supervised

Allenamento con dati etichettati (input-output noti). Esempi di algoritmi: ANN, SVM, Decision Tree, Random Forest.

Semi-supervised

Combinazione di pochi dati etichettati con molti dati non etichettati.

Un-supervised

Non ci sono etichette, l'algoritmo cerca pattern e relazioni nascoste (es. clustering). Algoritmi: PCA.

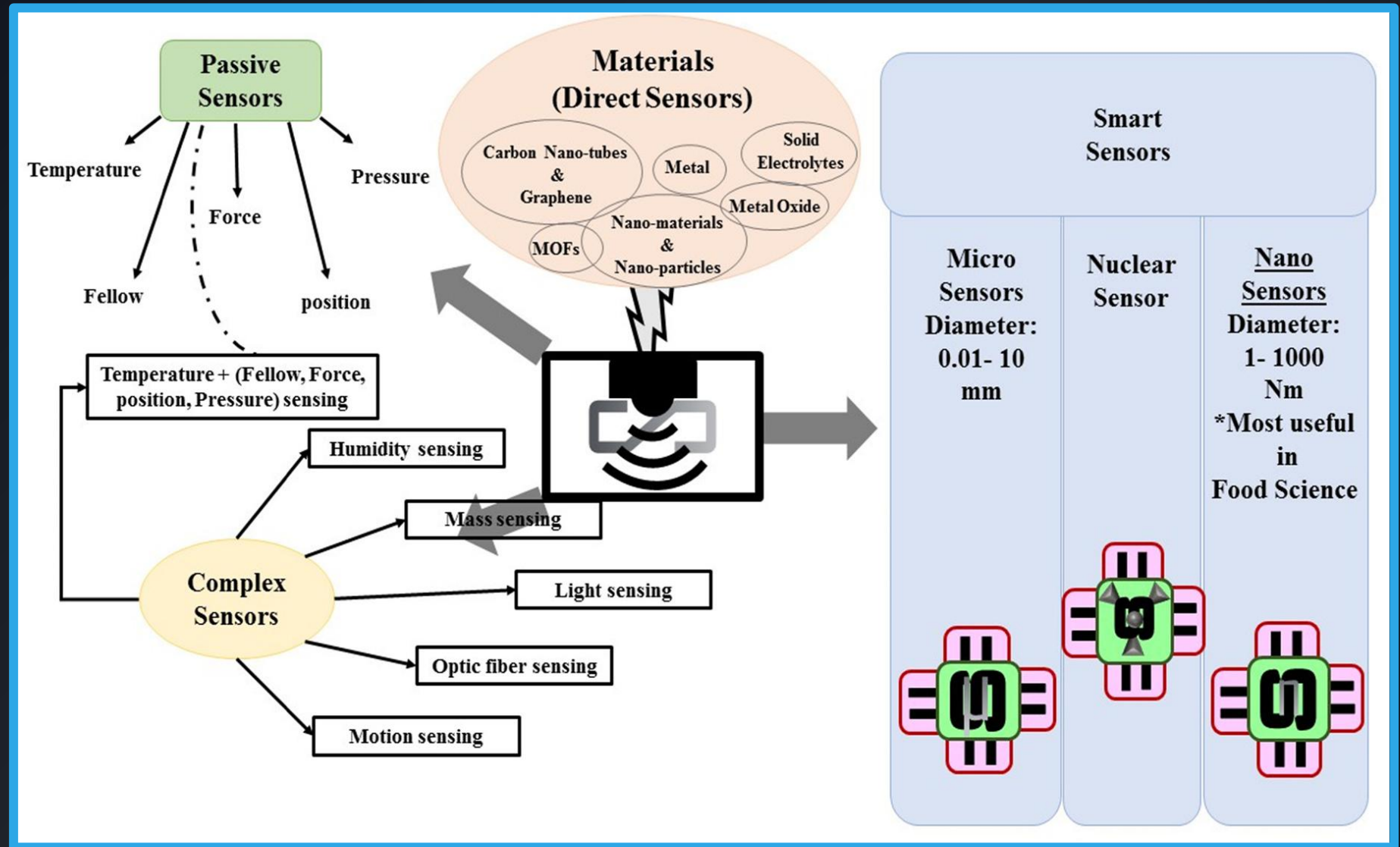
Reinforcement Learning

l'algoritmo "impara" tramite ricompense e penalità in un ambiente dinamico (es. robotica, giochi).

IoT e IoF: Connettività e dati “dal campo alla tavola”

Cos'è l'IoT?

Rete di sensori, dispositivi e software che trasmettono informazioni in tempo reale



Metodi Chemiometrici a supporto dell'Analisi Alimentare

Cosa sono i metodi chemiometrici?

- ❑ tecniche statistiche e matematiche applicate a dati chimico-fisici, come quelli provenienti da spettroscopia, cromatografia o imaging iperspettrale.
- ❑ spesso, queste tecniche includono strumenti di analisi multivariata come la PCA o la LDA

Perché sostituiscono la valutazione sensoriale?

- ❑ Riducono soggettività dell'analisi sensoriale "umana".
- ❑ Forniscono informazioni più dettagliate e quantitative su parametri come aroma, colore, stabilità.

Vantaggi

- ❑ Alta **precisione** e **ripetibilità**.
- ❑ **Integrazione con ML** (spesso in unsupervised analysis, es. PCA) per semplificare e "pulire" i dati, riducendo il rumore e le variabili ridondanti.

Pre-processing →
semplificazione dei dati →
input ottimizzato per
Machine Learning

Analisi della qualità e selezione dei prodotti

Sorting e ispezione con visione artificiale

- **TOMRA Sorting Solutions:** azienda leader che utilizza sensori e AI per l'ispezione automatica e la separazione di frutta, verdura e altri alimenti.
- **Greefa:** sviluppa sistemi di classificazione automatica per frutta e ortaggi, sfruttando telecamere ad alta risoluzione e algoritmi di riconoscimento dei difetti.



Previsione della domanda e ottimizzazione dell'inventario



Case study su supermercati e piattaforme di e-commerce

- **Morrisons & Blue Yonder:** la catena di supermercati britannici Morrisons ha collaborato con Blue Yonder (società di soluzioni AI) per implementare un sistema di previsione della domanda e ottimizzazione dell'inventario, che ha permesso di ridurre gli sprechi e evitare rotture di stock.
- **Amazon:** utilizza sistemi predittivi proprietari per anticipare la domanda di prodotti alimentari, gestire le scorte nei propri centri di distribuzione e pianificare le spedizioni Prime Now.

Sicurezza alimentare e tracciabilità

- **IBM Food Trust:** piattaforma blockchain che consente la tracciabilità di prodotti alimentari lungo l'intera filiera (adottata da Walmart, Carrefour, Nestlé, ecc.). Integrando AI e big data è possibile verificare in tempo quasi reale la provenienza e l'integrità dei prodotti.
- **Maersk & TradeLens:** pur essendo focalizzate principalmente sulla logistica dei container, hanno sperimentato soluzioni di tracciabilità end-to-end applicate al trasporto di prodotti alimentari a temperatura controllata.



Creazione di nuove ricette e Ricerca & Sviluppo

IBM Chef Watson: progetto sperimentale di IBM Watson che utilizzava un enorme database di ricette e principi di “food pairing” per proporre abbinamenti inediti di ingredienti e ricette innovative (alcuni ristoranti e chef hanno collaborato per testarne i risultati).

Supporto alla logistica e alla distribuzione

- **DHL e route optimization:** DHL impiega algoritmi di intelligenza artificiale per pianificare le rotte di consegna, ridurre i costi di carburante e migliorare i tempi di consegna, specialmente nel trasporto di merci deperibili
- **Ocado Technology:** la piattaforma di e-grocery inglese Ocado adotta sistemi AI per gestire gli ordini online, ottimizzare l'organizzazione dei magazzini automatizzati (con robot su griglia) e ridurre gli errori di picking.



Classificazione del profilo di tostatura con E-Nose e Deep Learning

Contesto

- Il grado di tostatura ha un impatto diretto sulla qualità sensoriale del caffè
- La valutazione tradizionale, affidata ad assaggiatori esperti, è costosa, poco scalabile e soggetta a errori umani.

Obiettivo dello Studio

Utilizzare un naso elettronico e una rete neurale (deep learning) per:

- Rilevare il profilo aromatico del caffè durante la tostatura
- Classificare automaticamente i campioni in base al grado di tostatura (da *light* a *dark roast*)

Perché è innovativo?

- Porta oggettività e automazione in un processo ancora soggettivo.
- Rende il controllo qualità replicabile, scalabile e preciso, con analisi in tempo reale del profilo aromatico

Dalla tostatura del caffè all'IA

PREPARAZIONE DEL CAMPIONE

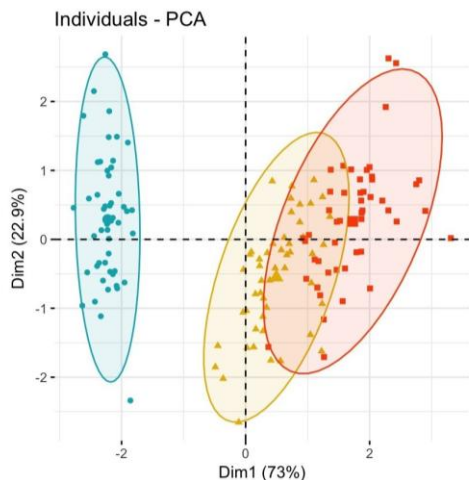
- 2 kg di caffè Robusta crudo, proveniente da Indonesia (Umidità iniziale: 12.9%)
- Divisi in 5 gruppi e tostati a temperature crescenti: 185°C – 195°C – 205°C – 215°C – 225°C → da light roast a dark roast
- Ogni campione è stato testato 50 volte per robustezza statistica

SENSORI E-NOSE (ARRAY TGS)

- 6 sensori MOS TGS: TGS 2620 (solventi/alcol); TGS 2612 (gas combustione); TGS 2611 (metano); TGS 2602 (composti organici volatili); TGS 2600 (odori di cottura/tabacco); TGS 826 (ammoniaca/ammine)
- Pre-riscaldamento 30 min prima della raccolta dati

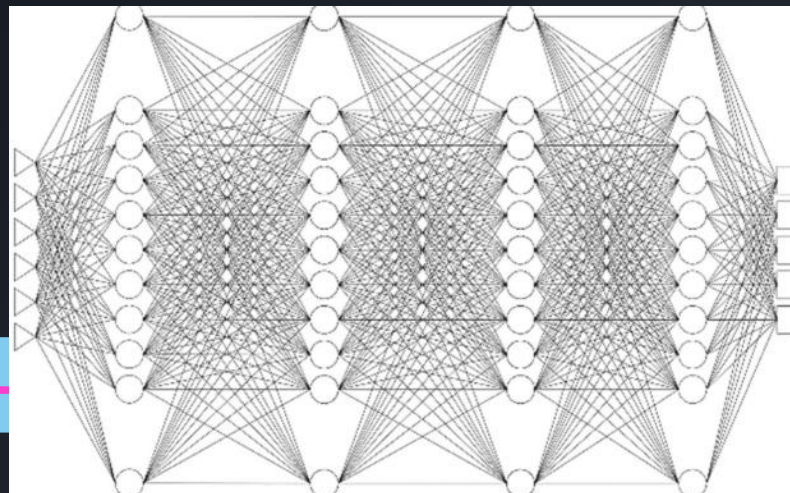
ALGORITMO DI CLASSIFICAZIONE

- Rete neurale profonda a 2 layer nascosti
- Tecniche:
 - i. Pearson Correlation per selezione sensori rilevanti
 - ii. Funzioni di attivazione: tansig, logsig, linear
 - iii. Ottimizzazione: algoritmo Adam
 - iv. Cross-validation a 10 fold
- Parametri ottimali: 91 neuroni, 3 layer, learning rate = 0.001

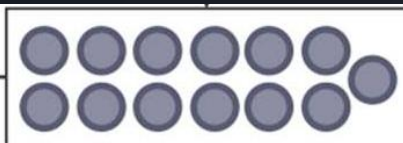


PATTERN RECOGNITION

Algoritmo di classificazione



PCA

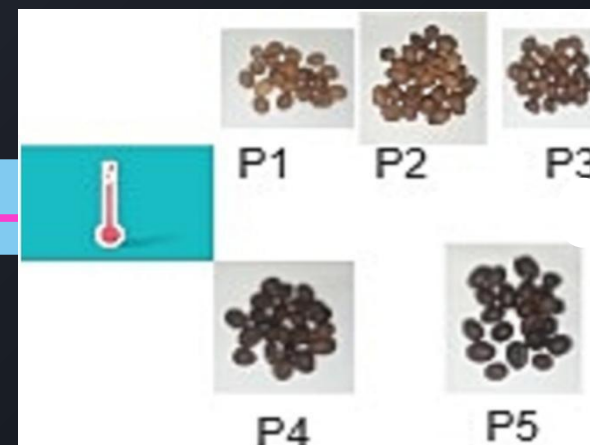


GAS SENSOR ARRAY

Gas Sensor Array

- TGS822
- TGS2611
- MQ4
- MQ136
- TGS2620
- MICS5524
- TGS2602
- Mq2
- MQ5
- TGS2600
- TGS826
- TGS2610
- TGS813

Naso elettronico

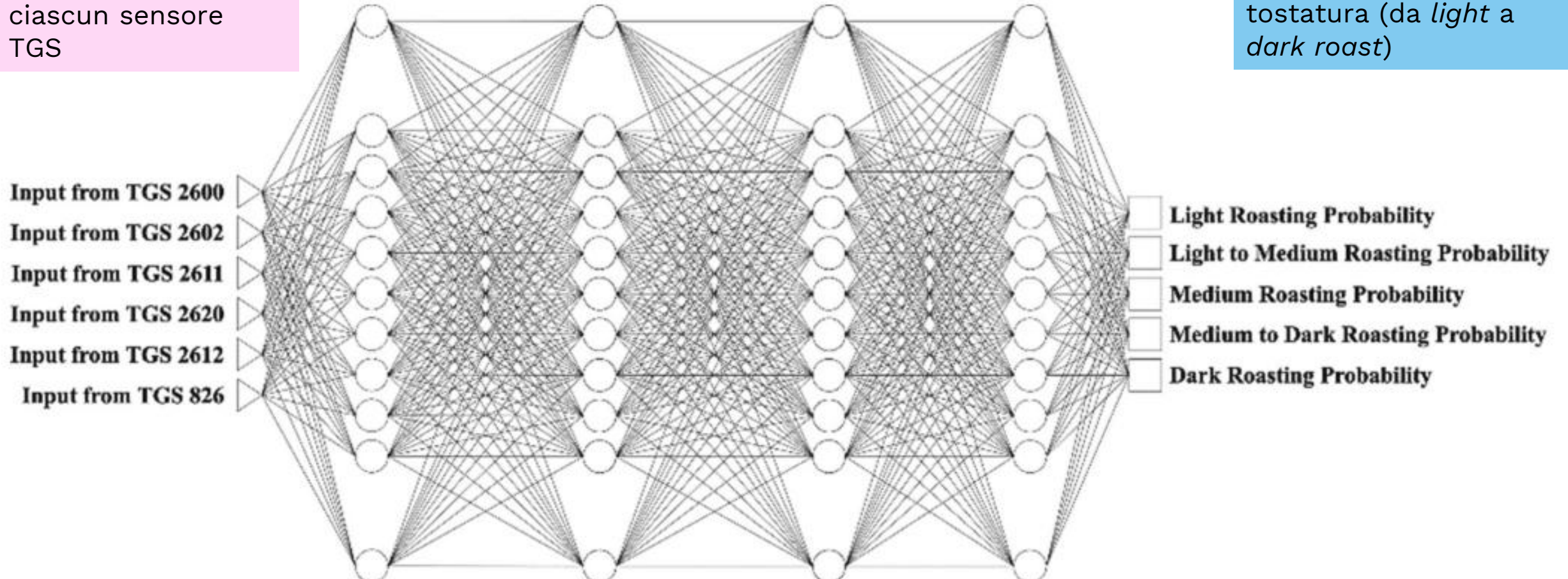


chicchi di caffè Robusta tostati
a 5 livelli di temperatura

Struttura della rete neurale

In **input**: segnali provenienti da ciascun sensore TGS

In **output**: probabilità per ciascun livello di tostatura (da *light* a *dark roast*)



Architettura della rete neurale usata per classificare il grado di tostatura dei campioni sulla base delle risposte dei sensori TGS.



CREAZIONE DI DATASET E LABELING PER ADDESTRAMENTO AI

SCATTARE FOTO

Richieste **MIGLIAIA**
di foto, quanto più
eterogenee possibili

ETICHETTARLE MANUALMENTE

Operazione **TEDIOSA**
che richiede molto
tempo e molto
lavoro manuale

SEGMENTARLE MANUALMENTE

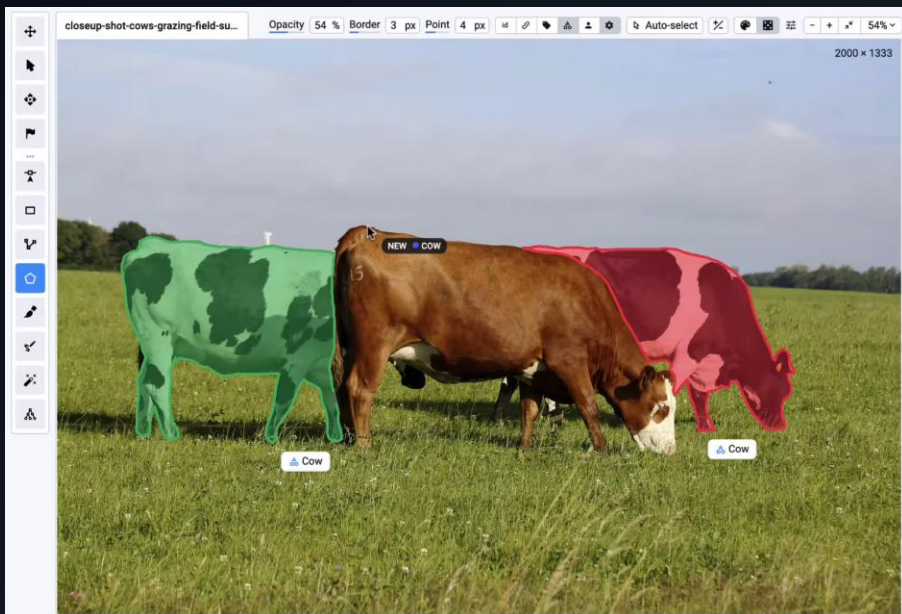
Operazione **TEDIOSA**
che richiede molto
tempo e molto
lavoro manuale

TEST DELLA FATTIBILITA'

Le foto scattate
potrebbero essere
non buone per
l'addestramento

TRAINING E DEPLOYMENT

Richiede hardware molto
potente e costoso, oltre a
tempi variabili dipendenti
dall'hardware



LABELING MANUALE

Segmentazione e/o bounding boxes

A seconda dell'algoritmo da utilizzare, le immagini devono essere etichettate secondo il formato richiesto (**YOLO per le bounding box, SAM, Detectron, YOLACT per la segmentazione**).

L'etichettatura manuale di decine di migliaia di immagini è lunga, costosa e laboriosa.

- Molti set di dati disponibili sono già etichettati ma, a parte casi specifici, sono “**generici**” e **non specializzati**.
- Inoltre, può essere **difficile acquisire fisicamente** le immagini per la costruzione del set di dati.



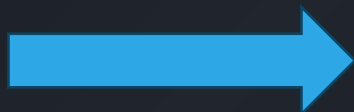
Fine-tuning di modelli pre-addestrati con dataset sintetici

1. Modelli pre-addestrati

- I modelli di ultima generazione sono dotati di architetture personalizzate e parametri pre-addestrati.
- L'addestramento da zero richiede grandi quantità di dati per ottenere prestazioni simili ed elevate potenze hardware.

2. Fine-tuning di modelli pre-addestrati

- Il fine-tuning regola i modelli pre-addestrati per compiti specifici utilizzando **nuovi dati**.
- La differenza fondamentale nella messa a punto è che si parte da parametri e pesi esistenti, a differenza dell'inizializzazione casuale dell'addestramento ex novo.



Generazione di dataset sintetici

- IA generativa (GAN, Stable Diffusion, ecc.)
- Generazione 3D procedurale (rendering 3D di scene controllate per ground truth)



IA GENERATIVA

IA generativa “pura” (es. GAN, Stable-Diffusion)

```
def _create_leaf_prompt(self, species):  
    """Create an optimized prompt for photorealistic leaf generation"""  
    prompts = [  
        f"Ultra detailed macro photograph of a single {species} leaf, natural lighting, photorealistic, 8k, professional photography, on white background",  
        f"Extremely detailed {species} leaf specimen, scientific illustration style, isolated, studio lighting, photorealistic",  
        f"Photorealistic {species} leaf with visible venation and texture, natural colors, botanical specimen, high resolution"  
    ]  
    return np.random.choice(prompts)
```



IA generativa di tipo “transfer style” (es. StyleGAN3)



- Utile per phenotyping, possibilità di avere ground truth per etichette
- **Non** ci sono informazioni spaziali (es. profondità spaziale)

GENERAZIONE 3D PROCEDURALE

$$\frac{dL_i}{dt} = \alpha * L_i * \left(1 - \frac{L_i}{L_{max}}\right)$$

→ Internode

$$\frac{dI}{dt} = c_I * \alpha * L_i * \left(1 - \frac{L_i}{L_{max}}\right) - k_i * I$$

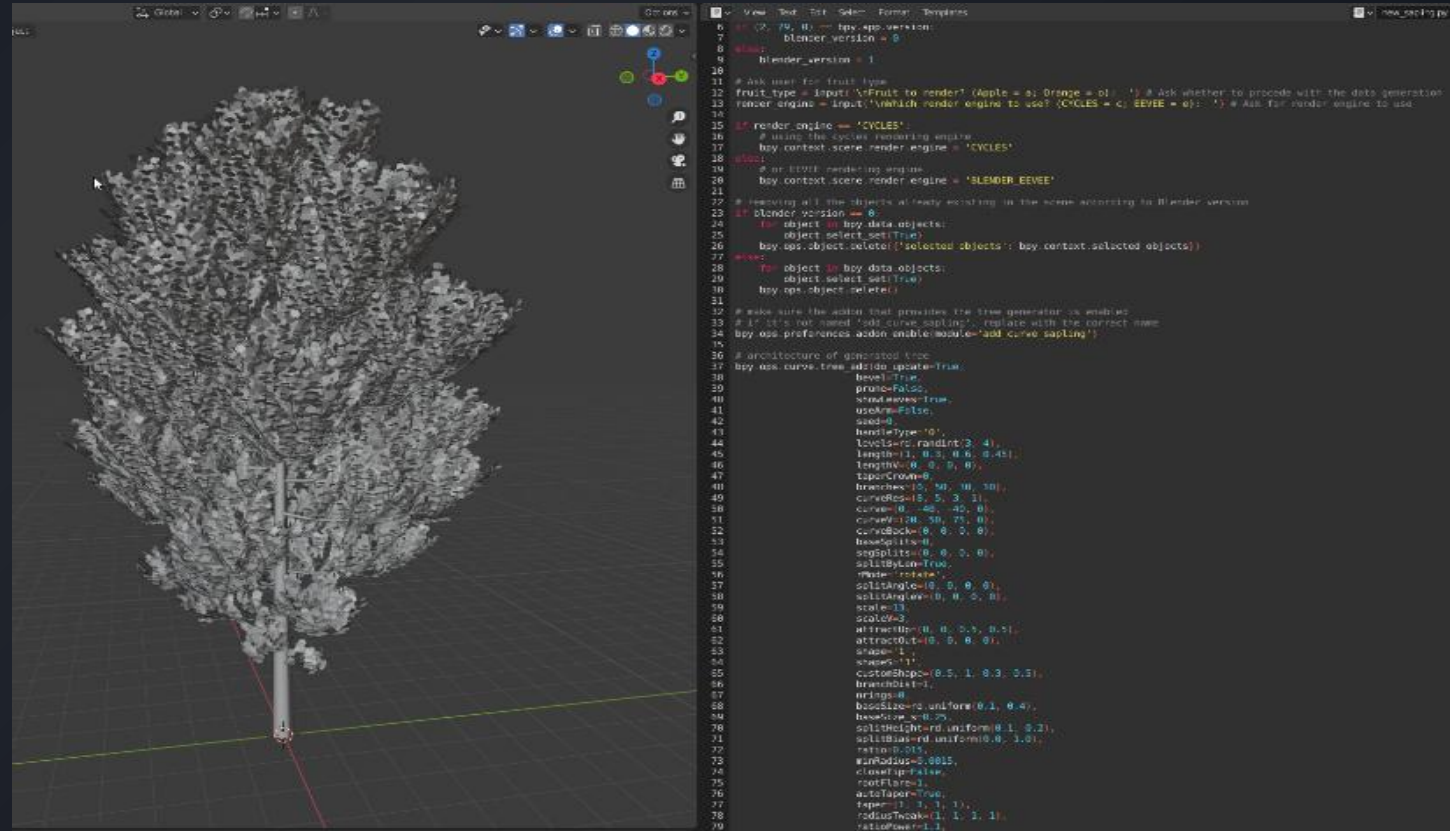
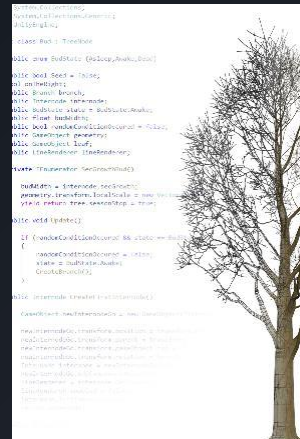
→ Branch

$$L_{max} = 1 - \frac{\sum L_i}{H_{max}} - 0.2 * \frac{I}{I_{max}}$$

→ Internode

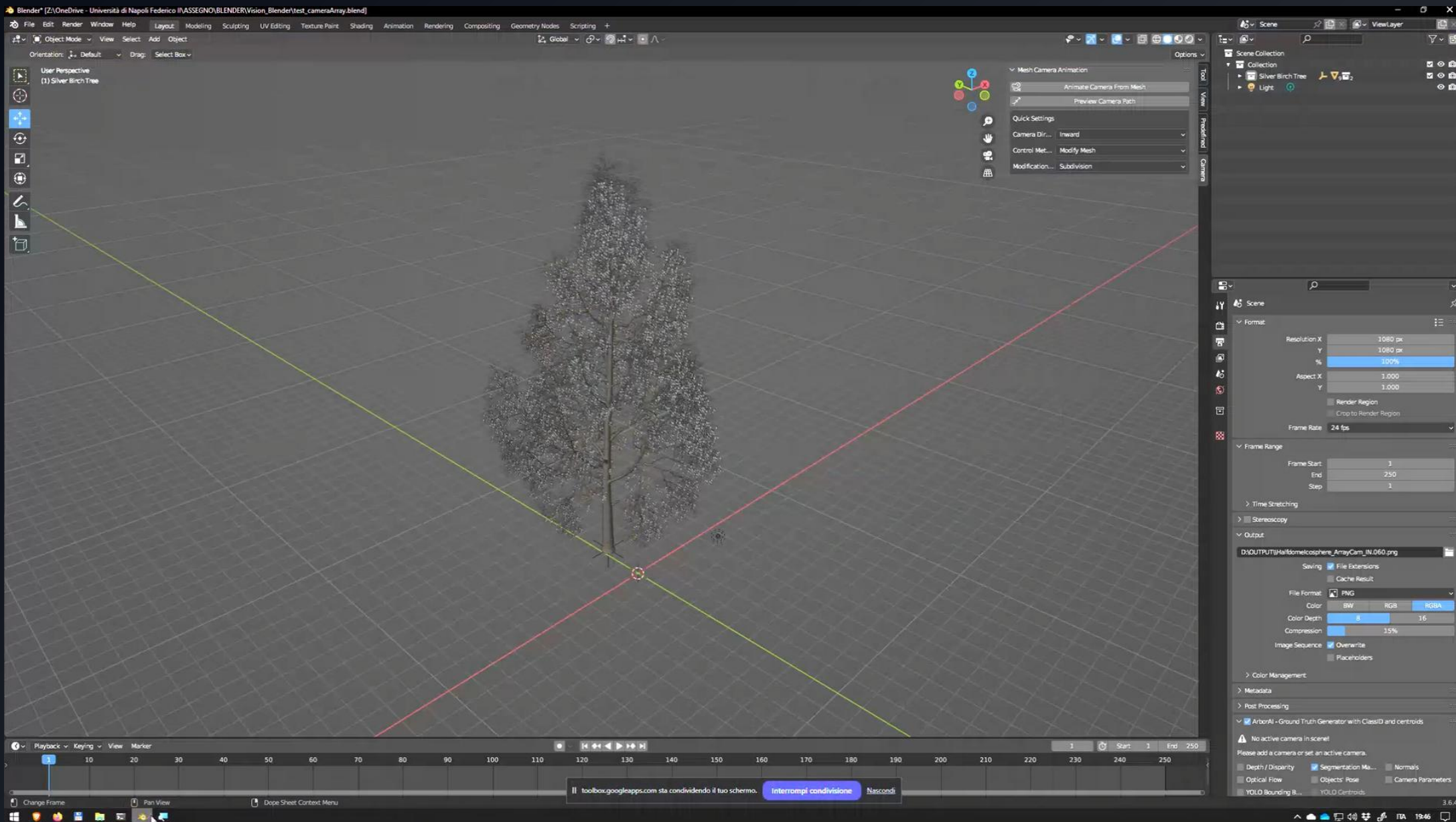
$$\frac{dD_i}{dt} = d_{actual} * D_i * \left(1 - \frac{D_i}{D_{max}}\right) * k_i$$

→ Internode

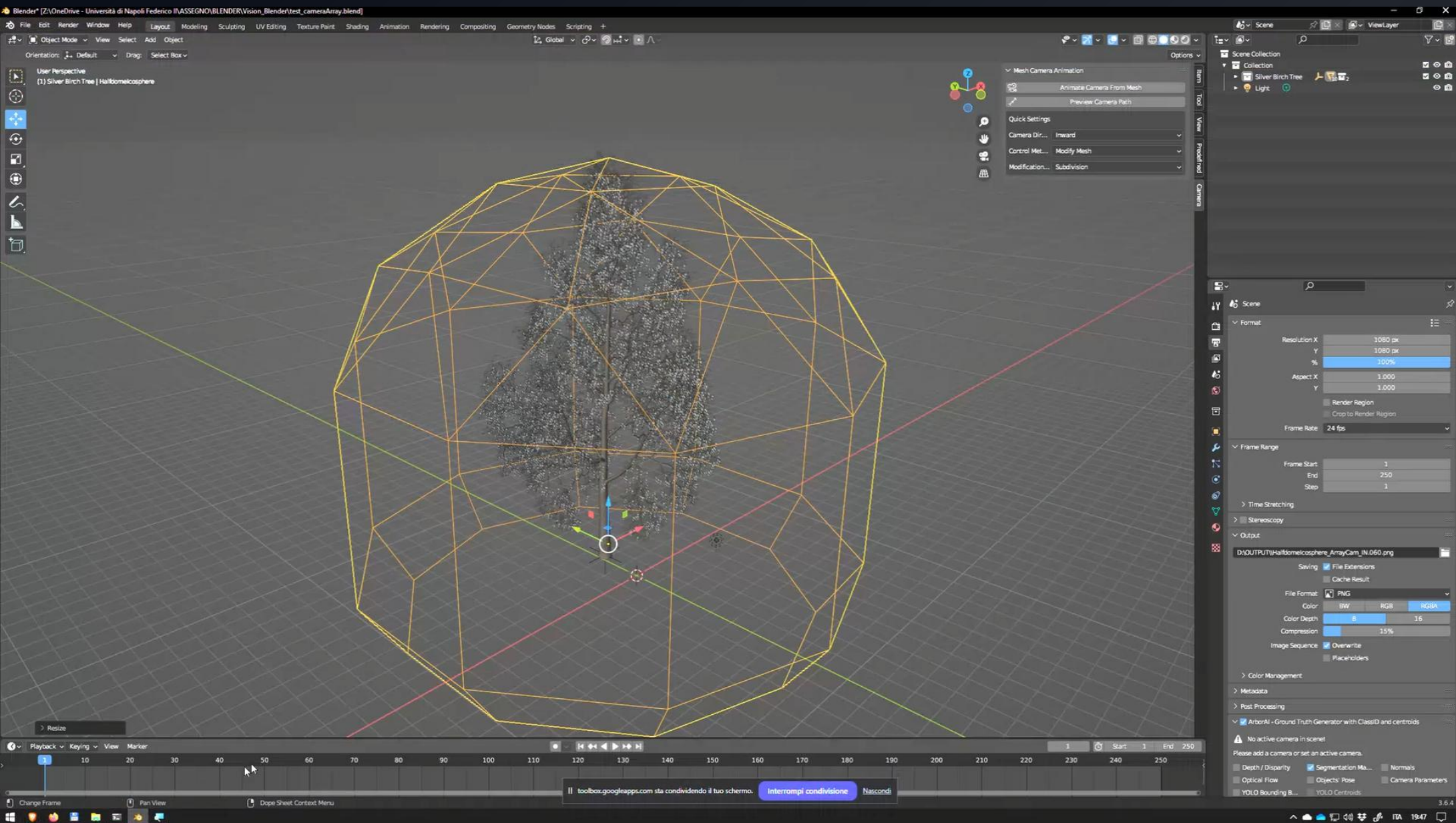


- Script Blender in Python per generazione automatica di alberi.
- Selta dei parametri di crescita dell'albero (angolo di ramificazione, gravitropismo, ecc.) con valori impostati o casuali.
- Scelta del tipo di frutto da modellare(mela, arancia, ecc.).
- Scelta del tipo di foglia.
- Scelta della quantità di frutti e di foglie.
- ...

Permette di avere il pieno controllo della scena, generando migliaia di immagini eterogenee (cambi di luce, di inquadratura, di sfondo) con annesse etichette, ground truth, segmentazioni, mappe di profondità, ecc...

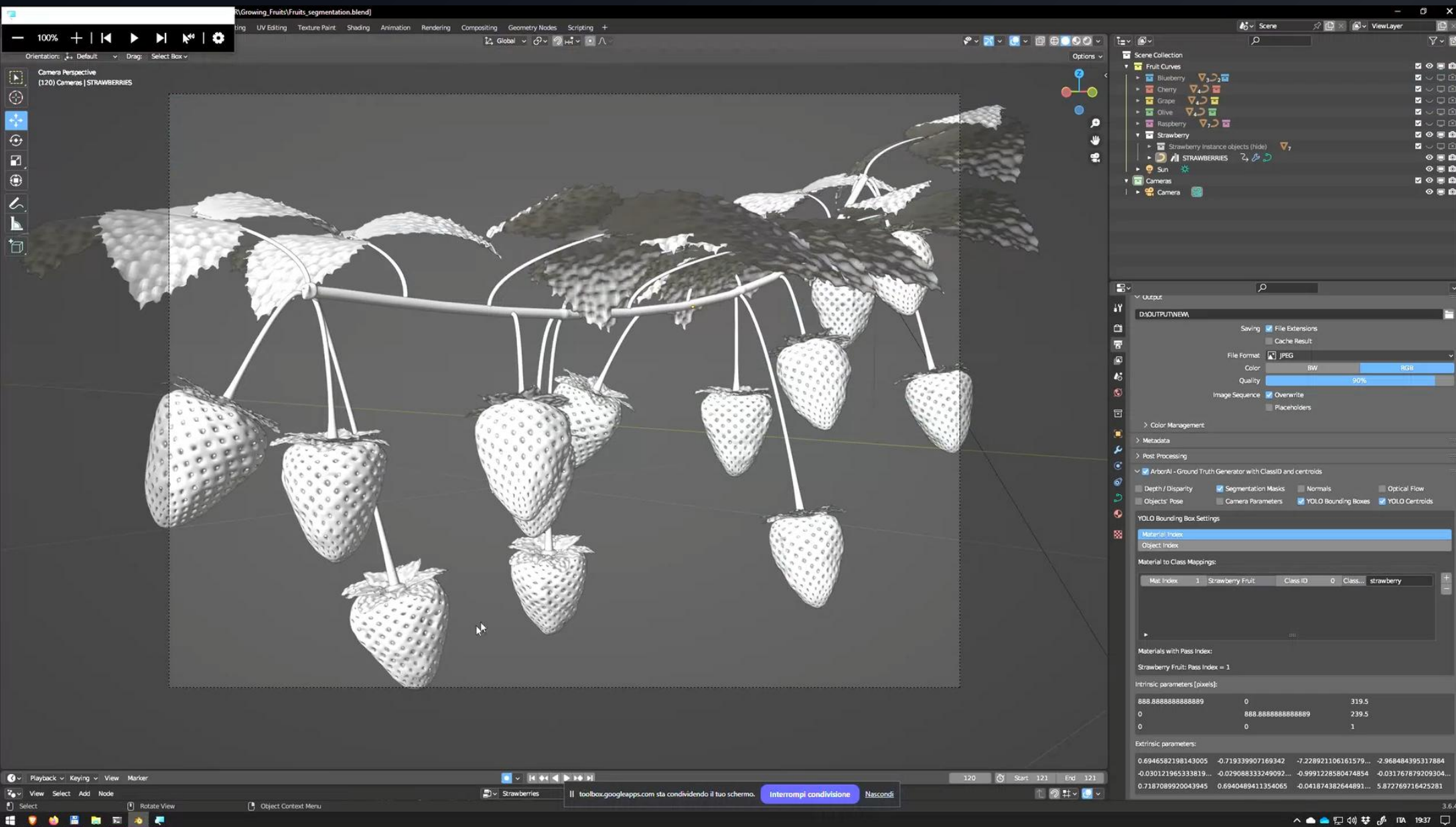


- Creazione forme predefinite per “avvolgere” il soggetto



- Creazione forme predefinite per “avvolgere” il soggetto

- Animazione camera lungo le facce per creare foto da migliaia di angolazioni diverse



- Creazione forme predefinite per “avvolgere” il soggetto

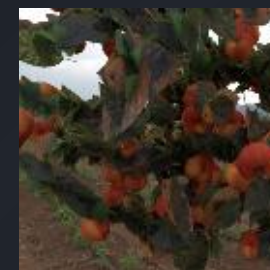
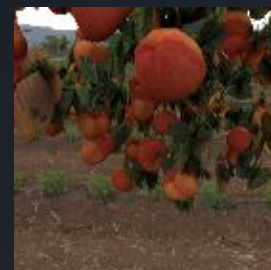
- Animazione camera lungo le facce per creare foto da migliaia di angolazioni diverse

- Creazione dei ground truth (segmentazione, YOLO bounding boxes, ecc.)

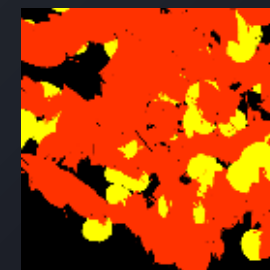
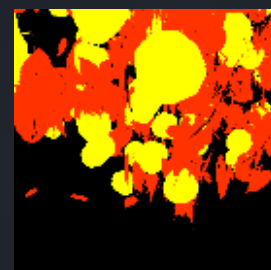
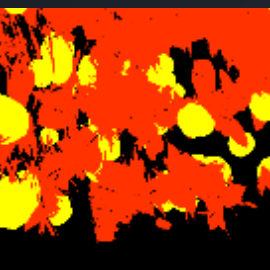
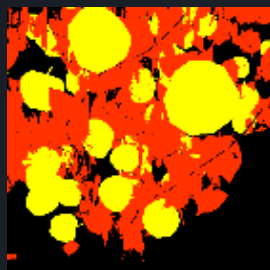


OUTPUT

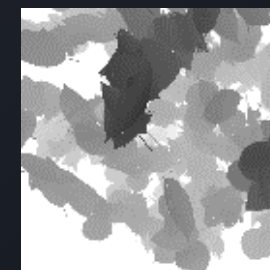
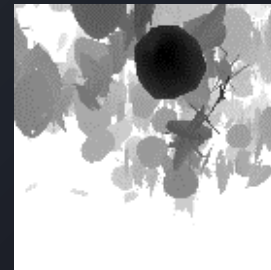
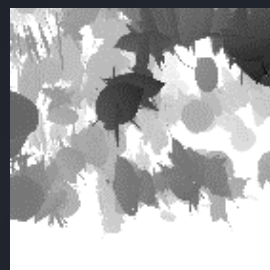
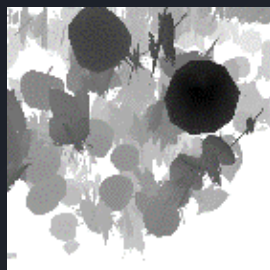
Immagini
sintetiche



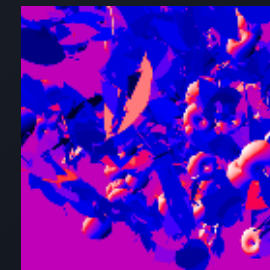
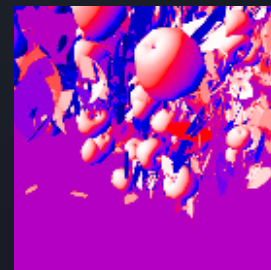
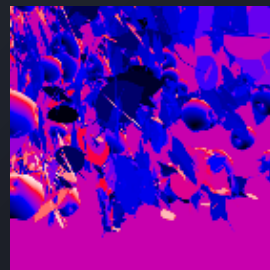
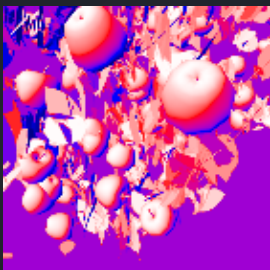
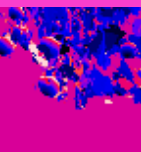
Mappe di
segmentazione



Mappe di
profondità



Mappe delle
normali

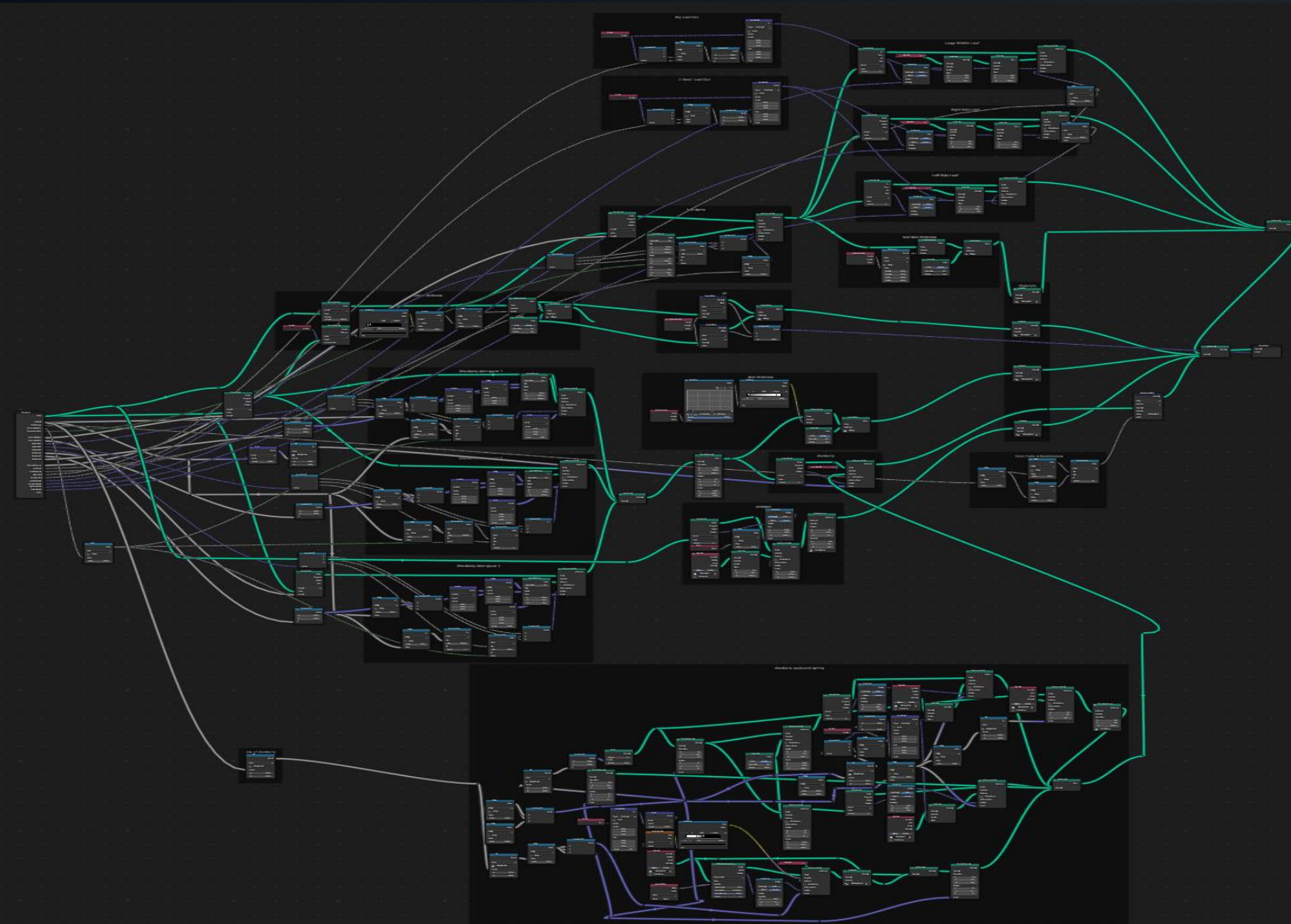


Struttura dei file di output

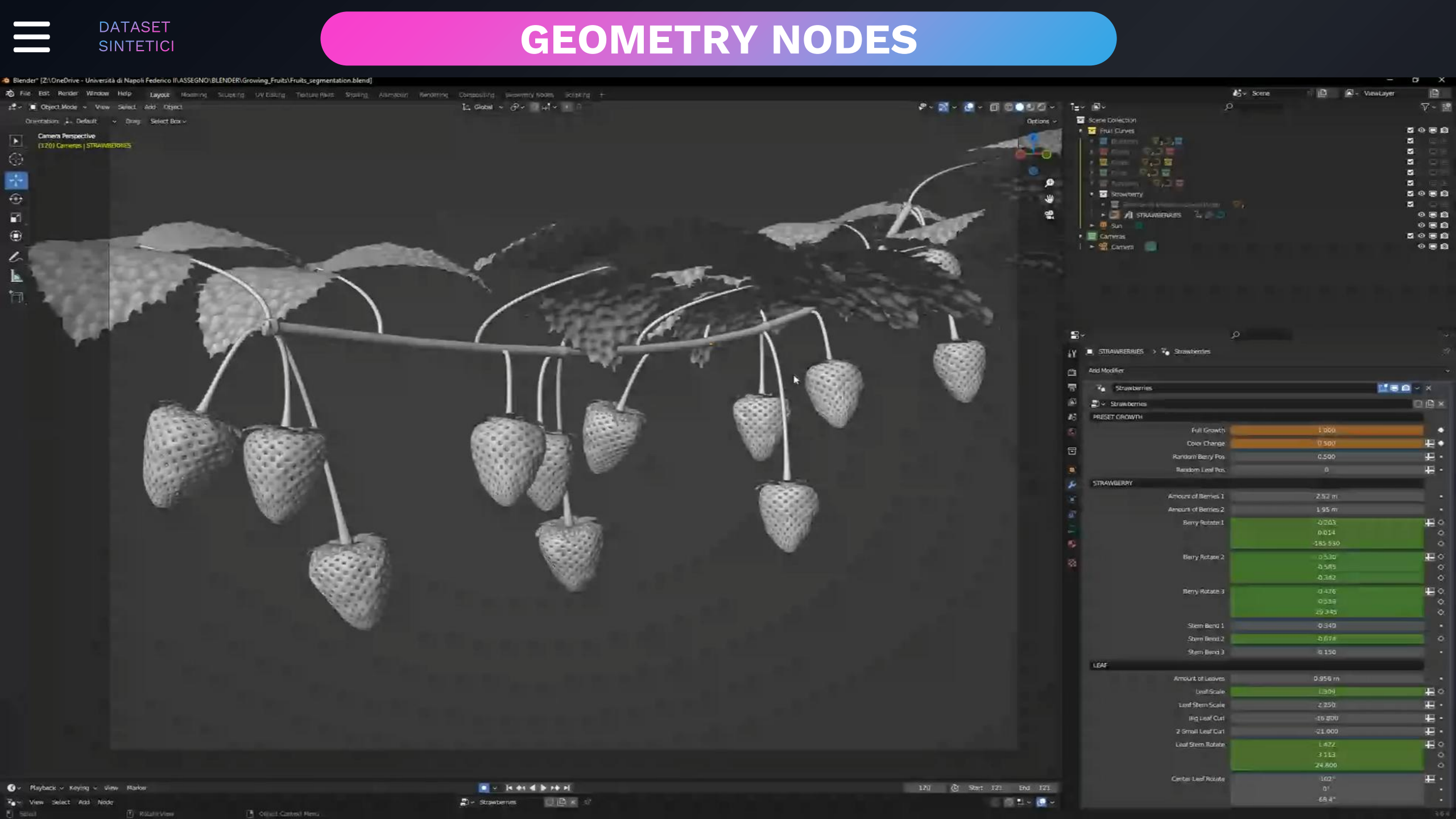
```
output_dir
- images
  - image0000
  - ...
- segmentation_map
  - image0000
  - ...
- depth_map
  - image0000
  - ...
- normal_map
  - image0000
  - ...

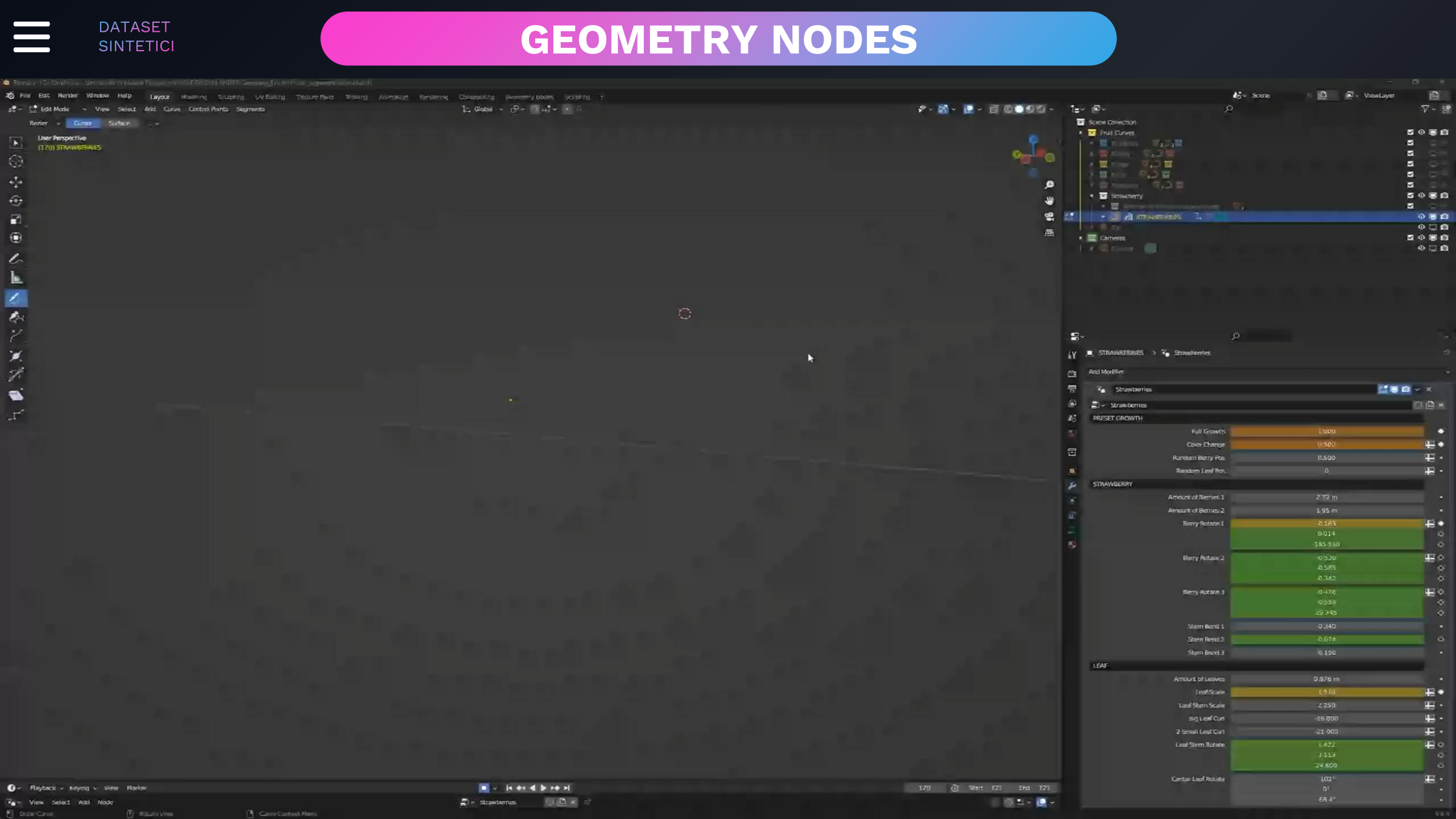
- image0000.txt
- image0000_centroids.txt
- classes.txt
```

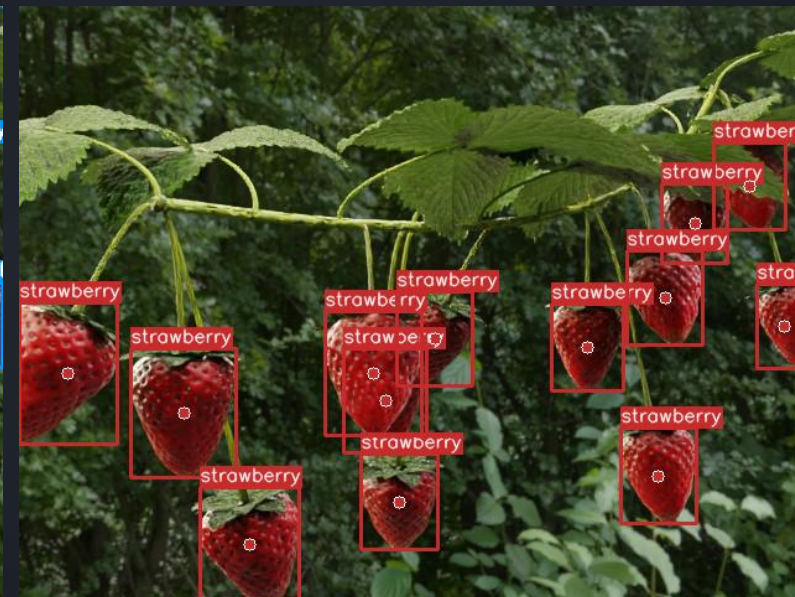
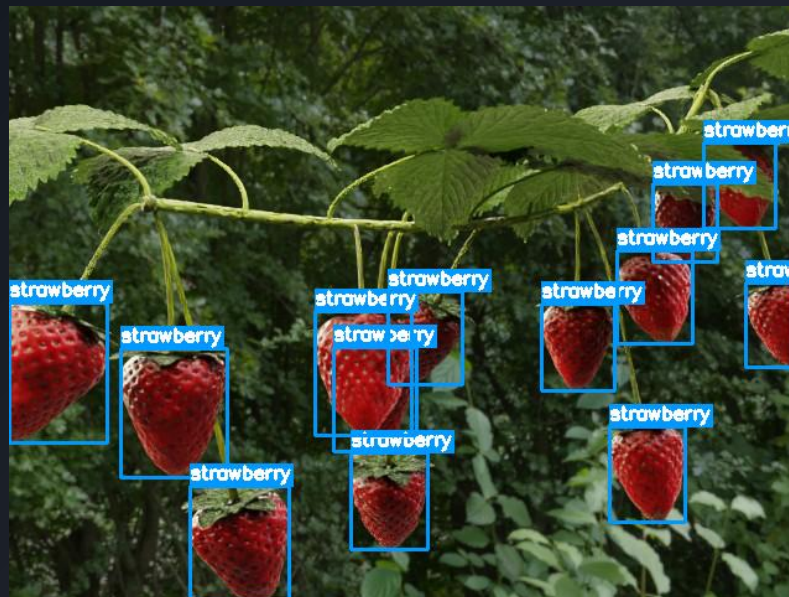
GEOMETRY NODES



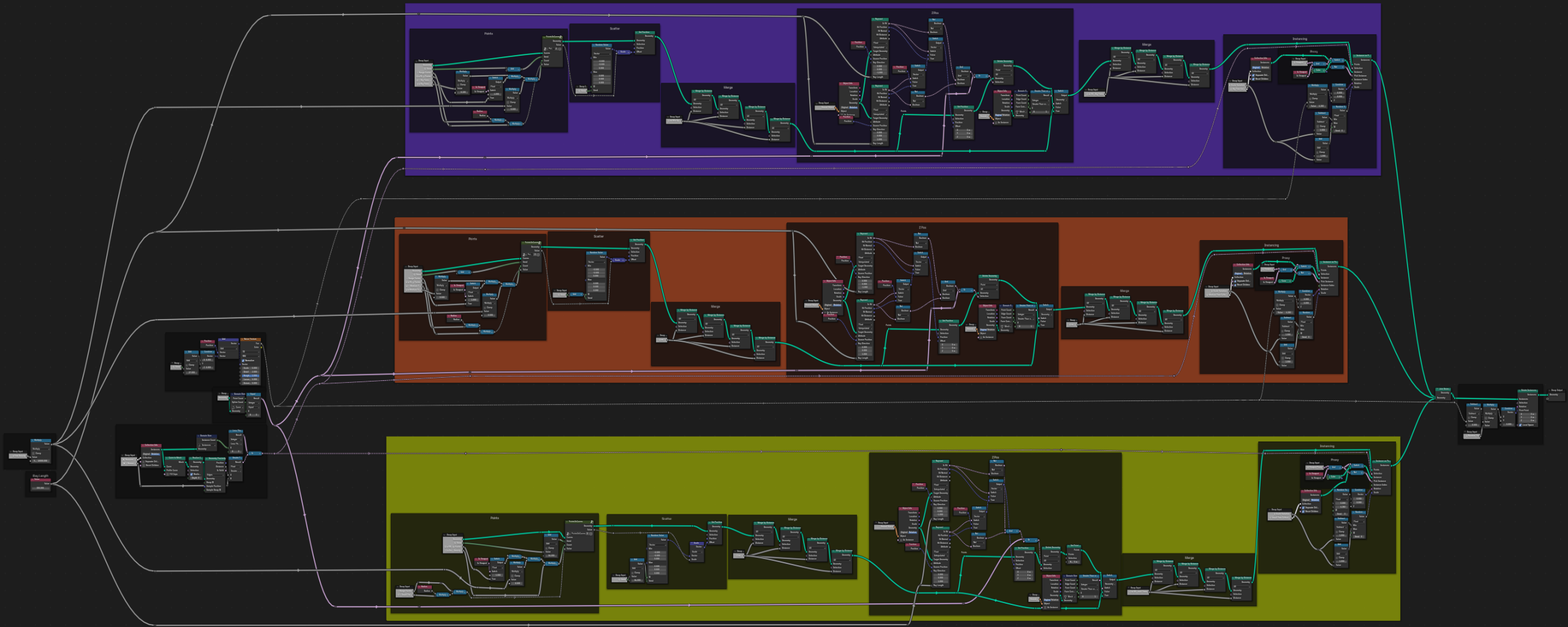
**Creazione procedurale e
in tempo reale di oggetti
in base a delle relazioni
(nodi) che richiamano le
leggi matematiche-
biologiche di crescita**

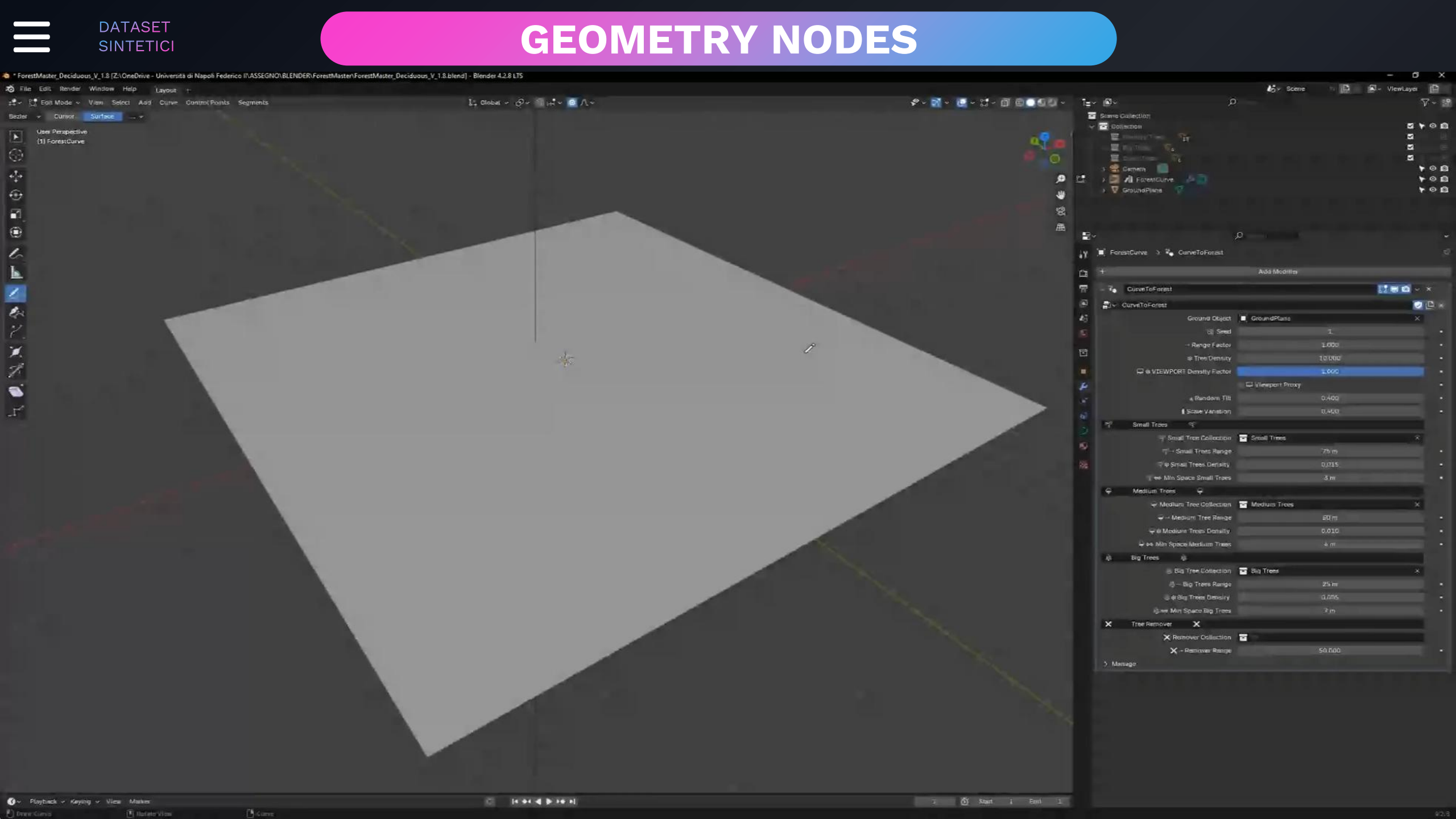






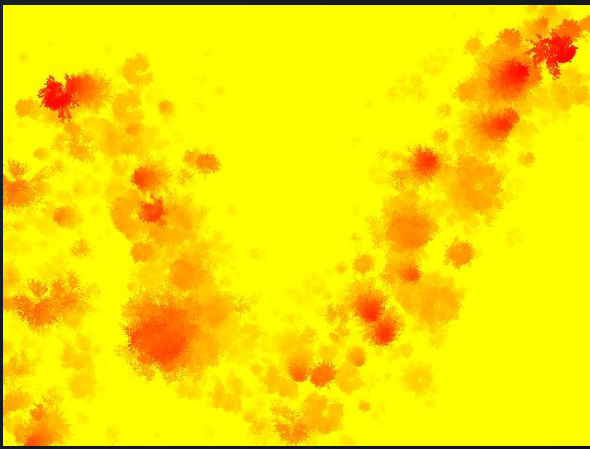
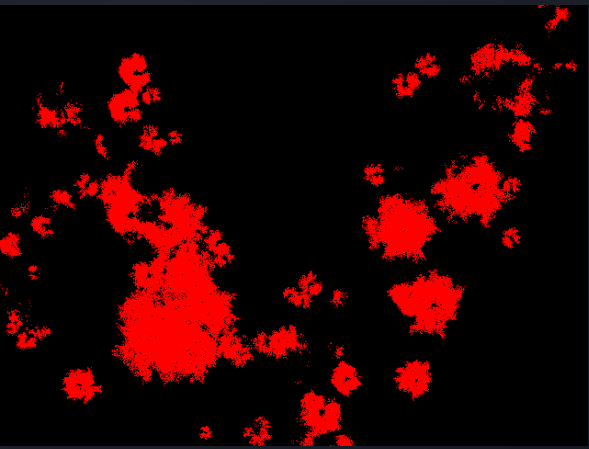
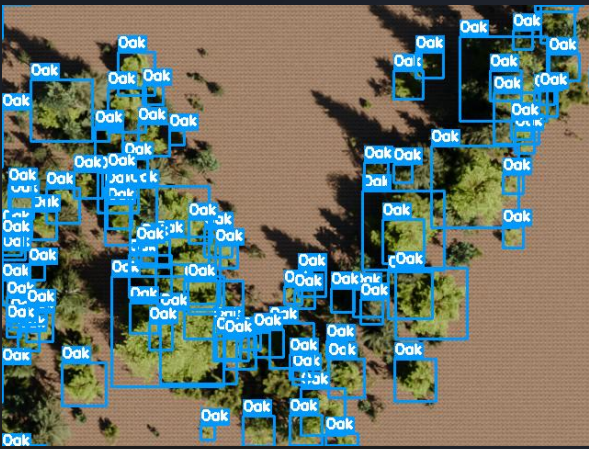
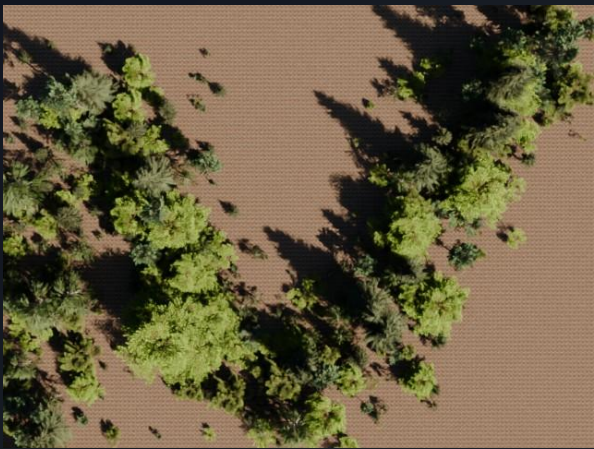
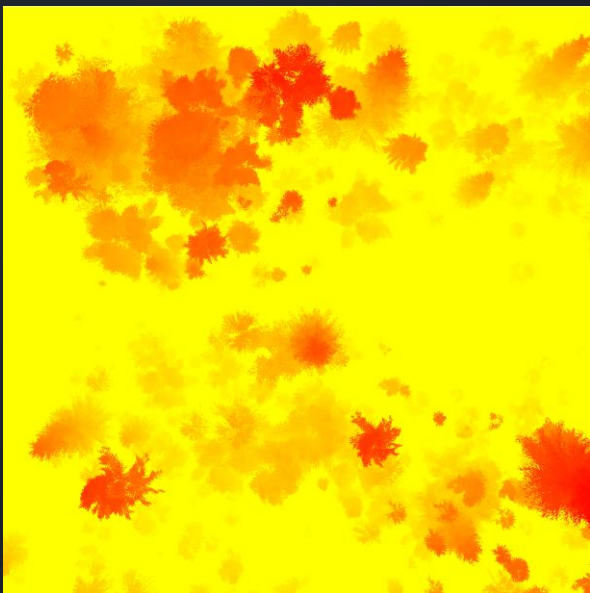
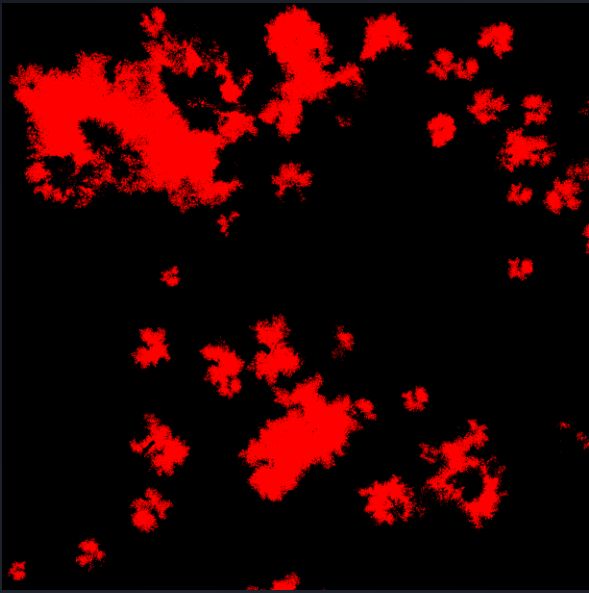
GEOMETRY NODES







GEOMETRY NODES



Math Expression

Result

lerp(0.5, (a² + 2ab) // π, b)

ab π

a

0.000

b



SymPy

[Main Page](#) [Features](#) [Download](#) [Documentation](#) [Support](#) [Development](#) [Roadmap](#) [Donate](#)

About

SymPy is a Python library for symbolic mathematics. It aims to become a full-featured computer algebra system (CAS) while keeping the code as simple as possible in order to be comprehensible and easily extensible. SymPy is written entirely in Python.

[Get started with the tutorial](#) [Download Now](#)

GRAZIE!

1. **Shumailov, I. et al. AI models collapse when trained on recursively generated data. *Nature* 631, 755–759 (2024).**
2. Afonso, M. & Giufrida, V. Editorial: Synthetic data for computer vision in agriculture. *Frontiers in Plant Science* 14, (2023).
3. Voetman, R., van Meekeren, A., Aghaei, M. & Dijkstra, K. Using Diffusion Models for Dataset Generation: Prompt Engineering vs. Fine-Tuning. in *Computer Analysis of Images and Patterns* (eds. Tsapatsoulis, N. et al.) 142–152 (Springer Nature Switzerland, Cham, 2023). doi:10.1007/978-3-031-44237-7_14.
4. Grondin, V., Pomerleau, F. & Giguère, P. Training Deep Learning Algorithms on Synthetic Forest Images for Tree Detection. Preprint at <https://doi.org/10.48550/arXiv.2210.04104> (2022).
5. Barth, R., IJsselmuiden, J., Hemming, J. & Henten, E. J. V. Data synthesis methods for semantic segmentation in agriculture: A Capsicum annuum dataset. *Computers and Electronics in Agriculture* 144, 284–296 (2018).
6. Chen, R. J., Lu, M. Y., Chen, T. Y., Williamson, D. F. K. & Mahmood, F. Synthetic data in machine learning for medicine and healthcare. *Nat Biomed Eng* 5, 493–497 (2021).

PROSSIMO APPUNTAMENTO

7 MAGGIO 2025 – *Emanuele D’Anza, Giuseppe Terlizzi*