Grayscale Image Colorization Based on Semantic Segmentation

VIORICA, PUȘCAȘ
UNIVERSITATEA TRANSILVANIA DIN BRAȘOV
Facultatea de Matematică și Informatică
Specializarea: Informatică aplicată
Email: viorica.puscas@student.unitbv.ro

Abstract

Colorizing grayscale images is a prominent challenge in Computer Vision, with numerous deep learning approaches being prevalent in the literature for this task. This paper introduces a multi-head fully convolutional network architecture, taking grayscale images as input and outputting the chromaticity information for colorization and segmentation masks for semantic segmentation. The model is inspired by the one proposed by Iizuka et al. in "Let there be Color!: Joint End-to-end Learning of Global and Local Image Priors for Automatic Image Colorization with Simultaneous Classification" (2016), but goes a step further by transitioning from iconic images to images with a number of different object categories. Preliminary empirical results, measured on both subjective and objective scales, demonstrate convincing colorization, without artifacts or color bleeding.

Keywords: *colorization, convolutional neural network, computer vision*

Introducere

Colorarea imaginilor în tonuri de gri este o problemă esențială în Computer Vision care a atras atenția a numeroși cercetători în ultimii ani, folosind în special metode de colorare bazate pe tehnici de deep learning. Imaginile color oferă o experiență vizuală mai plăcută și sunt folosite cu precădere în probleme din Computer Vision, precum clasificarea, detectarea de obiecte și segmentarea semantică a imaginilor. Astfel, problema colorării a fost studiată în detaliu prin prisma a mai multor aplicații, printre care se numără: restaurarea imaginilor istorice [1] [2] [3], colorarea imaginilor infraroșu [4] [5], colorarea cadrelor din animații etc.

Problema colorării imaginilor în tonuri de gri poate avea mai multe soluții corecte, întrucât un obiect poate fi colorat în mai multe moduri, fără a reduce din realismul imaginii obținute. Prin urmare, o soluție corectă este o colorare plauzibilă a unei imagini în tonuri de gri.

Punct de plecare

Începând cu anul 2012, rețelele neurale convoluționale se bucură de o mare popularitate în rândul cercetătorilor în ceea ce privește sarcinile din Computer Vision. Problema colorării automate a imaginilor în

tonuri de gri a fost de asemenea abordată cu ajutorul a astfel de rețele neurale, cu care s-au obținut rezultate remarcabile.

Modelul propus de Iizuka *et al.* [1] se distinge de alte abordări din literatura de specialitate bazate pe rețele neurale convoluționale prin introducerea clasificării ca sarcină auxiliară celei principale, de colorare.

Conform autorilor, rolul clasificării este de a corecta erori de colorare cauzate de lipsa unui context

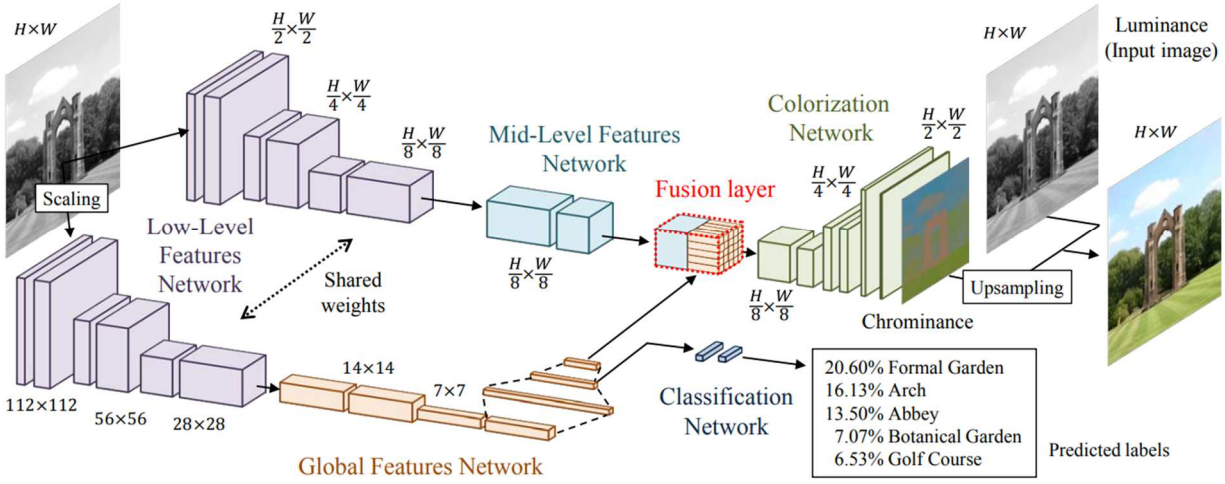


Fig. 1. Modelul propus de Iizuka et al. [1]

global al imaginii care să ghideze extragerea de trăsături prin intermediul straturilor convoluționale. Alegerea de a optimiza simultan pentru colorare și pentru clasificare este justificată în lucrare prin publicarea rezultatelor unui studiu conform căruia imaginile colorate astfel sunt considerate naturale în procent de 92.6%, în timp ce imaginile colorate cu un model antrenat fără partea de clasificare sunt considerate naturale în 69.8% din cazuri.

CIELAB

În literatura de specialitate, problema colorării este abordată cu precădere în spațiul de culoare CIELAB [1] [2] [3]. Acest spațiu de culoare este cunoscut pentru faptul că este uniform din punct de vedere perceptual, unde orice schimbare numerică corespunde unei schimbări similare în percepția umană asupra culorii. De asemenea, în CIELAB, informația legată de luminozitate este separată de informația cromatică. Astfel, problema colorării este simplificată întrucât este necesară prezicerea valorilor pentru doar două dintre cele trei canale, canalul L fiind echivalent cu imaginea în tonuri de gri care se dorește a fi colorată.

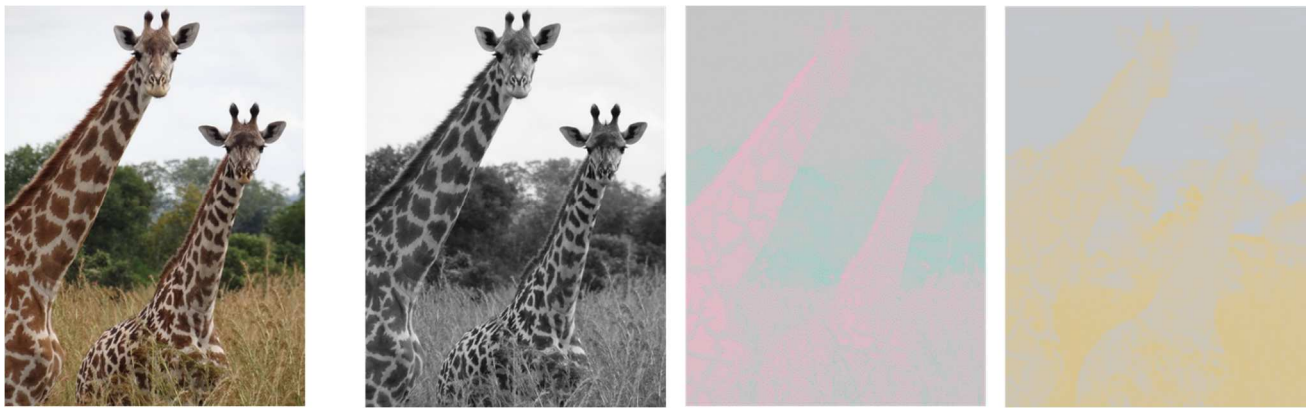


Fig. 2. De la stânga la dreapta: imaginea RGB; canalul L în spațiul de culoare CIELAB, echivalent luminozității; canalul a, reprezentând spectrul de culori de la verde la roșu; canalul b, reprezentând spectrul de culori de la albastru la galben

Autor: Viorica Pușcaș

Colorare bazată pe segmentare semantică

Prin introducerea clasificării ca sarcină auxiliară colorării, rezultatele modelului propus de Iizuka *et al.* [1] s-au dovedit a fi vizibil mai realiste decât cele obținute cu un model de tip encoder-decoder clasic. Alegerea de a folosi clasificarea pentru a veni în ajutorul colorării este intuitivă, întrucât încadrarea unei imagini într-o clasă presupune extragerea trăsăturilor astfel încât contextul global al imaginii să fie asociat clasei respective. Dată fiind natura acestei sarcini, imaginile prezente în seturile de date pentru clasificare sunt imagini iconice[1]. Pornind de la această observație, o întrebare care poate apărea este: Ce se întâmplă în cazul imaginilor care conțin mai multe obiecte din diferite categorii?

Segmentarea semantică este o altă sarcină din Computer Vision care diferă de clasificare prin faptul că presupune atribuirea unei clase fiecărui pixel dintr-o imagine, spre deosebire de atribuirea unei clase întregii imagini. Astfel, trecerea de la clasificare la segmentare semantică ca și sarcină auxiliară colorării este una naturală, care promite depășirea limitării impuse de clasificare de a colora doar imagini în tonuri de gri iconice. În urma unei analize a literaturii de specialitate am constatat faptul că nu există nicio lucrare publicată în care colorarea să fie ajutată prin optimizarea simultană pentru segmentare semantică.

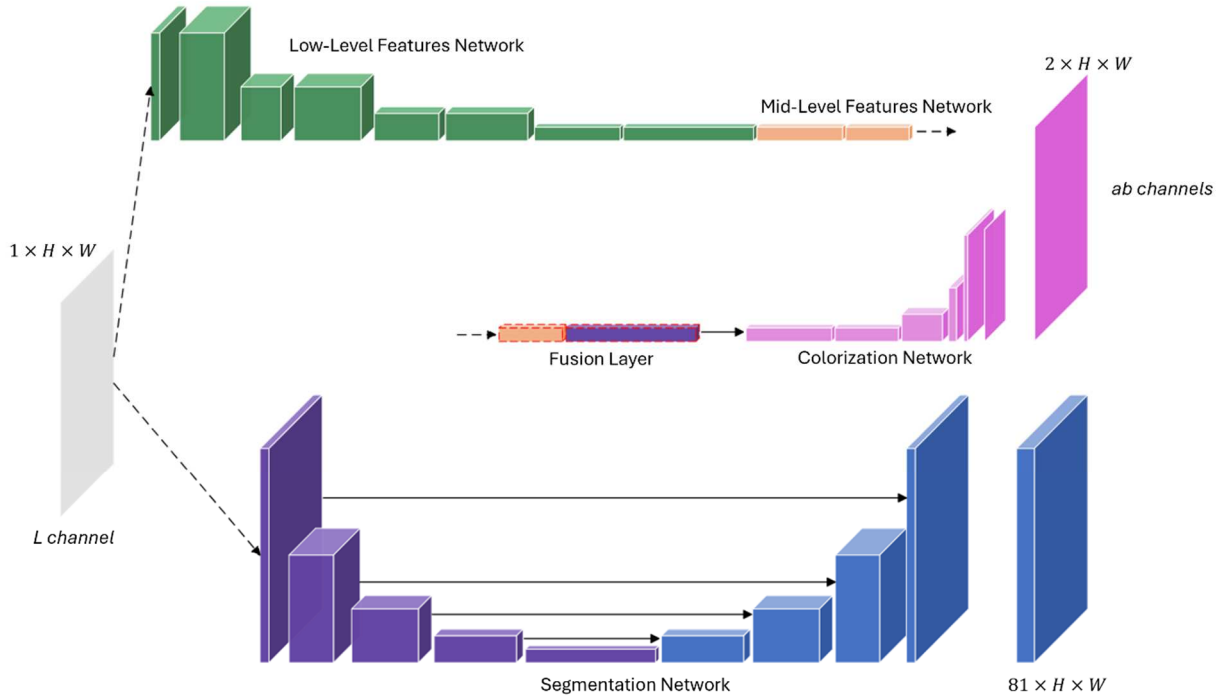


Fig. 3. UColNet v1

Autor: Viorica Pușcaș

UColNet v1 este primul model obținut în care clasificarea este înlocuită cu segmentarea semantică pentru îmbunătățirea colorării. El este inspirat de cel propus de Iizuka *et al.* [1] prin păstrarea a două ramuri specifice fiecărui proces, de colorare, respectiv segmentare semantică, și a stratului de fuziune introdus de către autori pentru combinarea trăsăturilor extrase de către cele două ramuri. Ramura de jos, care avea ca și scop final clasificarea, a fost înlocuită cu un model U-Net [6], iar fuziunea se realizează între cele două volume aflate la cea mai mică dimensiune spațială.

[1] Imagini cu un singur obiect, de obicei centrat.

Modelul a fost antrenat pe setul de date COCO [7] – subsetul *train*, 90% din imagini fiind folosite pentru antrenare și 10% pentru validare. Funcția de eroare folosită pentru colorare este Mean Squared Error, iar pentru segmentare am folosit Cross-Entropy și Dice înmulțite cu un coeficient adaptiv α pentru ca în timpul optimizării să se acorde aceeași importanță celor două sarcini. Ca optimizator s-a folosit SGD cu momentum 0.9, weight decay 1e-5 și o rată de învățare adaptivă, pornind de la 0.1 și fiind redusă la 90% din valoarea precedentă la fiecare 5 epoci.

În urma antrenării acestui model s-a constatat că este supraparametrizat, graficele funcțiilor de eroare pe setul de antrenare și pe cel de validare indicând fenomenul de overfitting.

UColNet v2 este primul model obținut prin încercarea de a simplifica modelul descris anterior. Acesta este un model U-Net în care separarea dintre cele două sarcini, cea de colorare și cea de segmentare semantică, este realizată prin ultimul strat convoluțional din decoder. S-a concluzionat devreme în procesul de antrenare faptul că această separare intervine prea târziu și că un singur strat convoluțional nu conține suficienți parametri pentru a modela rezultatul pentru fiecare sarcină în parte. Astfel, performanța acestui model este prea modestă pentru a se justifica introducerea lui detaliată în lucrare, dar consider că încercarea merită a fi menționată.

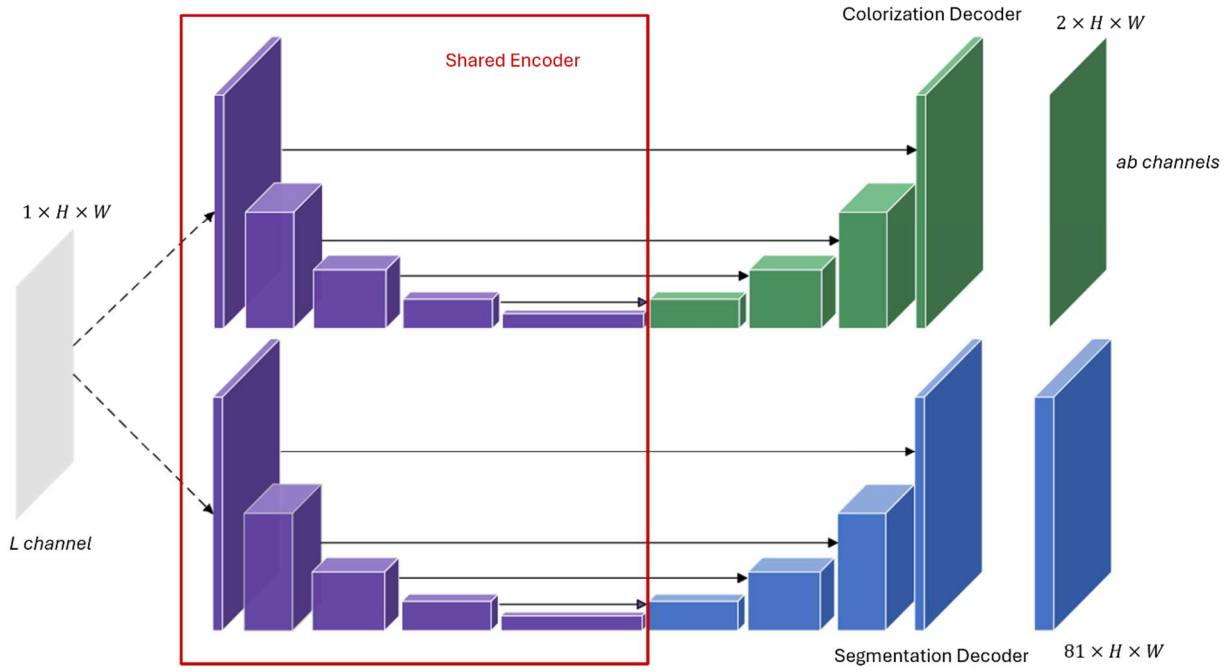


Fig. 4. UColNet v3
Autor: Viorica Pușcaș

UColNet v3 este cel mai bun model obținut prin simplificarea primei versiuni. Este un model în stil U-Net cu două ramuri care împart același encoder pentru extragerea de trăsături, urmat de câte un decoder pentru fiecare dintre cele două sarcini ale modelului: colorare și segmentare semantică.

Acest model a fost antrenat în aceleași condiții descrise mai sus, cu excepția termenului pentru regularizare. Motivul pentru care am renunțat la regularizarea L2 în antrenarea modelului UColNet v3 a fost pentru a studia capacitatea de generalizare a acestuia în urma modificărilor aduse. După antrenare s-a constatat faptul că fenomenul de overfitting apare în continuare, dar mult mai târziu decât în cazul primului model. Efectul regularizării asupra UColNet v3 nu a fost încă studiat.

Rezultate și evaluare

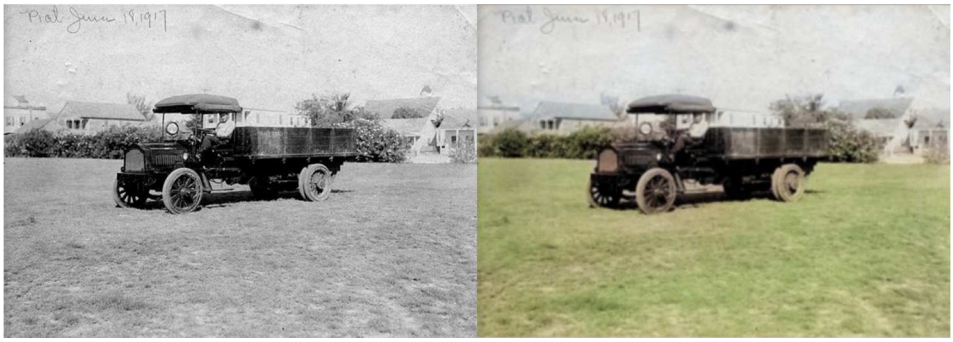
Problema colorării este cunoscută ca fiind greu de evaluat obiectiv, întrucât nu există o metrică perceptuală, care să corespundă cu percepția umană asupra unei imagini. În literatură, modelele de colorare propuse sunt evaluate obiectiv cu ajutorul unor metrici bazate pe diferența dintre imaginea color inițială și imaginea recolorată, a căror rezultate penalizează cazurile în care imaginea recolorată, deși realistă, este diferită de original. O altă metodă de evaluare întâlnită este cea subiectivă, realizată de oameni, a căror rezultate, deși utile pentru a demonstra calitatea modelului propus, nu sunt reproductibile și pot fi influențate de un grad mare de subiectivitate.

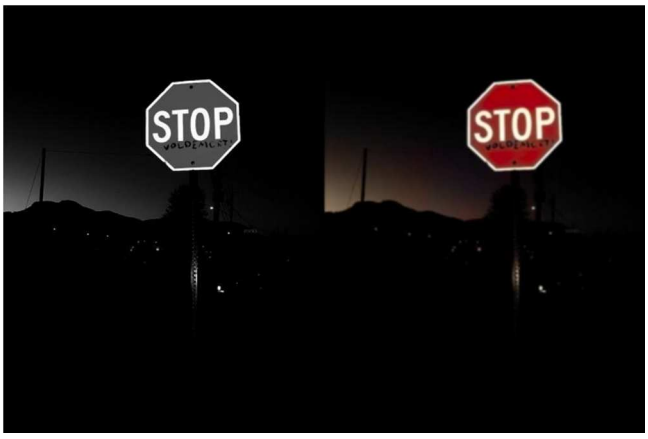
În scopul evaluării modelelor pentru colorare propuse, am introdus o nouă metodă de cuantificare a calității rezultatului, bazată pe detectare de obiecte. Folosind un model YOLOv8 [8] preantrenat, imaginile din setul de testare (subsetul *val* al setului de date COCO) sunt date ca intrare modelului pentru detectare de obiecte, rezultând numărul de obiecte identificate corect. Numărul de obiecte identificate corect după recolorare este obținut cu același model YOLOv8 cu care inferența se realizează pe aceleași imagini din setul de testare, transformate în tonuri de gri și colorate cu modelul ce se dorește a fi evaluat. Metoda de evaluare introdusă se bazează pe faptul că o colorare realistă permite modelului YOLOv8 să recunoască un procent similar de obiecte ca și în imaginile color inițiale.

În tabelul de mai jos am raportat atât valorile obținute în urma evaluării bazată pe detectarea de obiecte, cât și valorea metricii Root Mean Square Error.

Model	RMSE	Nr. obiecte GT	Nr. obiecte recunoscute GT	Procent GT	Nr. obiecte recunoscute colorare	Procent colorare
UColNet v1	18.45	36781	17554	47.73%	13546	36.83%
UColNet v3	18.33			47.73%	13459	36.59%

Imaginile color de mai jos au fost colorate cu UColNet v3, acestea provin din setul de date folosit pentru testare.




STOP
STOP

Concluzii

În cadrul acestei lucrări am propus două modele pentru colorare inspirate din [1] care se folosesc de segmentare semantică ca și sarcină auxiliară colorării, metodă care nu este întâlnită în literatura de specialitate. În urma unei evaluări subiective, se poate constata că imaginile colorate cu ajutorul celui mai bun model obținut sunt realiste, lipsite de artefacte sau transfer de culoare între obiecte. De asemenea, am introdus o nouă metrică pentru evaluarea obiectivă a colorării, bazată pe detectare de obiecte. Această metrică poate fi considerată a fi mai bună decât metricile obiective folosite până în prezent în literatură întrucât nu penalizează colorările realiste, dar diferite de imaginile inițiale, precum se întâmplă în cazul metricilor bazate pe diferența între imagini.

La momentul scrierii acestei lucrări, cercetarea este încă în desfășurare. În ceea ce privește perspectivele de viitor ale cercetării, este cunoscut faptul că arhitectura U-Net nu mai este actuală și că există alte modele propuse în literatură cu ajutorul cărora se obțin rezultate mai bune în cazul segmentării semantice. De asemenea, funcția de eroare Mean Squared Error folosită pentru colorare este o funcție bazată pe diferența între imagini. O altă direcție în care se poate îndrepta cercetarea este studiul efectului pe care alte funcții de eroare folosite în literatură îl au asupra modelelor propuse bazate pe segmentare, culminând cu găsirea unei funcții de eroare perceptuale, care să fie echivalentă cu percepția umană asupra realismului unei imagini. Conform cunoștințelor mele actuale, o astfel de funcție nu a fost propusă în lucrările publicate care au ca subiect colorarea de imagini în tonuri de gri.

Bibliografie

1] S. Iizuka, E. Simo-Serra și H. Ishikawa, „Let there be color!: joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification,” *ACM Transactions on Graphics (ToG),* vol. 35, nr. 4, pp. 1-11, 2016.

2] R. Zhang, P. Isola și A. A. Efros, „Colorful image colorization,” în *Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14*, 2016.

3] G. Larsson, M. Maire și G. Shakhnarovich, „Learning representations for automatic colorization,” în *Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part IV 14*, 2016.

4] X. Kuang, J. Zhu, X. Sui, Y. Liu, C. Liu, Q. Chen și G. Gu, „Thermal infrared colorization via conditional generative adversarial network,” *Infrared Physics & Technology,* vol. 107, p. 103338, 2020.

5] P. L. Suárez, A. D. Sappa și B. X. Vintimilla, „Infrared image colorization based on a triplet DCGAN architecture,” în *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017.

6] O. Ronneberger, P. Fischer și T. Brox, „U-net: Convolutional networks for biomedical image segmentation,” în *Medical image computing and computer-assisted intervention-MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, 2015.

7] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár și C. L. Zitnick, „Microsoft COCO: Common Objects in Context,” în *Computer Vision-ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, 2014.

Ultralytics, „YOLOv8: A New State-of-the-Art Computer Vision Model," [Interactiv]. Available:
8] https://yolov8.com. [Accesat 18 mai 2024].