מצגת סיום — RealOrFake

פרויקט גמר | קורס מולטימדיה ולמידת מכונה | 2025–2026

8 שקפים · ~5 דקות

Made with GAMMA

זיהוי תמונות מזויפות בבינה מלאכותית — RealOrFake

על הפרויקט

הוא כלי מבוסס למידת מכונה לזיהוי אוטומטי של תמונות RealOrFake שנוצרו על ידי בינה מלאכותית. המערכת מנתחת כל תמונה תוך שברירי שמסבירה *למה* GradCAM שנייה, מספקת ציון ביטחון ומציגה מפת חום מלא בעברית ובאנגלית Web המודל הגיע למסקנתו — הכל בממשק.

הצוות

אדיר שלומוב ו**פלורה**

קורס: מולטימדיה ולמידת מכונה
שנת לימודים: 2025–2026

פרויקט גמר — זיהוי תמונות AI בזמן אמת

Made with GAMMA

האתגר: לא ניתן לדעת אם תמונה אמיתית

כלי AI כמו Midjourney ו-DALL·E 3 מייצרים תמונות פוטוריאליסטיות ברמה שלא ניתן להבחין בהן בעין בלתי מזוינת. הבעיה אינה תיאורטית — היא כבר כאן, ומשפיעה על מיליוני אנשים מדי יום.

90% טועים

מחקר 2024 מראה ש-90% מהאנשים **לא מבחינים** בין תמונה אמיתית לתמונת AI — גם כשמדובר בתמונות ברזולוציה גבוהה

פייק ניוז

תמונות מזויפות מופצות ברשתות חברתיות כדי **לעוות מציאות**, להשפיע על דעת קהל ולערער אמון בתקשורת

הונאות וזיוף

זהויות מזויפות, **ראיות מפוברקות** ופרופילים בדויים — כלים חזקים בידי גורמים זדוניים

פתרון אוטומטי עם 98.28% דיוק — RealOrFake

מה המערכת עושה?

אפליקציית Web מלאה (בעברית ואנגלית) המזהה תמונות AI בזמן אמת. המשתמש מעלה תמונה, והמערכת מחזירה תשובה תוך ~66 מילישניות — כולל ציון ביטחון ומפת חום GradCAM שמסבירה את ההחלטה.

- העלאת קובץ תמונה
- הזנת URL ישיר
- צילום במצלמה
- ניתוח סרטון קצר

המספרים מדברים

98.28%
דיוק המודל

~66ms
זמן עיבוד לתמונה

56K
תמונות באימון

המודל מבוסס על **EfficientNet-B0** — ארכיטקטורה יעילה שמשלבת דיוק גבוה עם מהירות עיבוד מצוינת, מה שהופך אותה לאידיאלית לשימוש בזמן אמת.

Made with GAMMA

Defactify Dataset — 56,000 תמונות אמיתיות ו-AI

הדאטהסט שלנו בנוי משני מקורות משלימים: תמונות צילום אמיתיות ממאגר MS COCO המוכר, ותמונות שנוצרו על ידי ארבעה גנרטורי AI מובילים. חוסר האיזון בין המחלקות (יחס של 1:2.5) טופל באמצעות **class weights** בתהליך האימון.

📷 Real — MS COCO

16,000 תמונות צילום יומיומי אותנטי

- אנשים, טבע, אוכל, ספורט
- מגוון תנאי תאורה ורקעים
- תמונות ממקורות ציבוריים

🤖 AI Generated

40,000 תמונות מארבעה גנרטורים מובילים

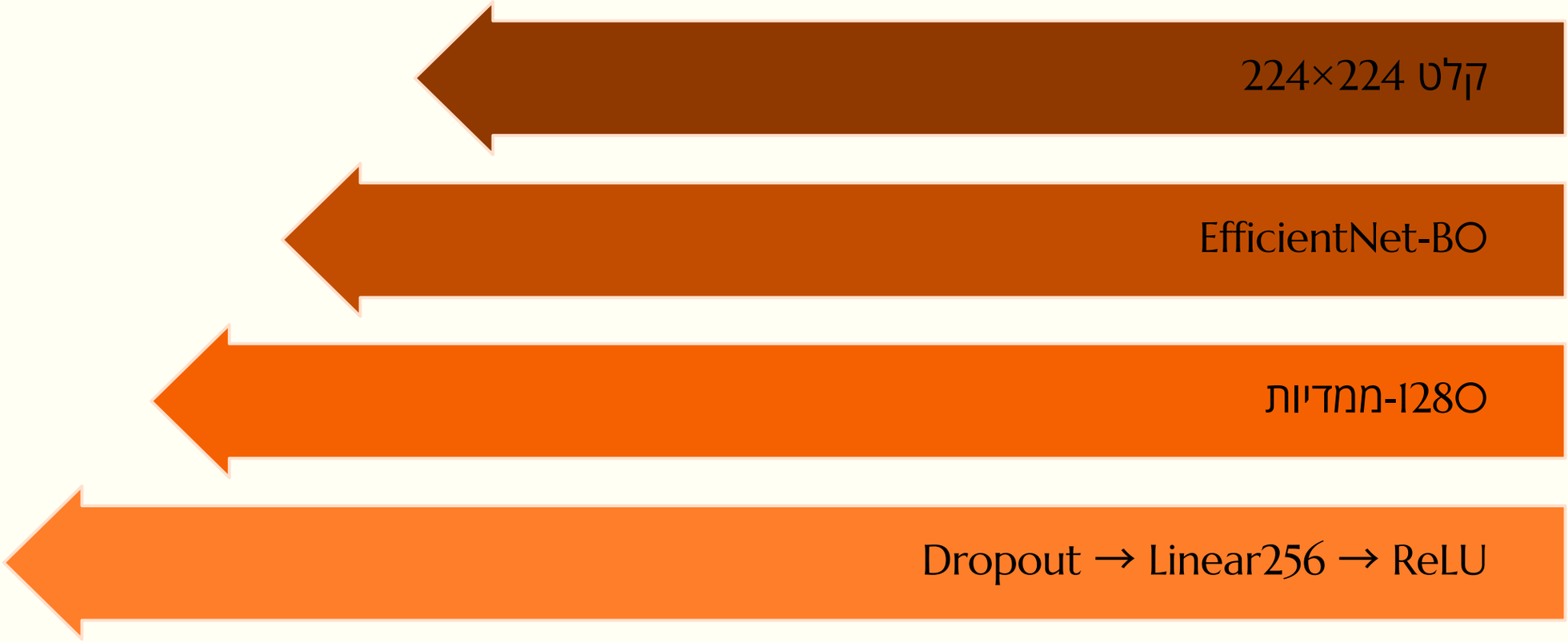
- Midjourney v6
- DALL·E 3 (OpenAI)
- Stable Diffusion XL
- Stable Diffusion 3

ⓘ חוסר האיזון בדאטה (1:2.5) פתרנו באמצעות **class weights** — משקל גבוה יותר לתמונות Real כדי שהמודל לא יטה לכיוון AI.

Made with GAMMA

Transfer Learning — EfficientNet-B0 + ImageNet

במקום לאמן רשת נוירונים מאפס, השתמשנו ב-**Transfer Learning**: לקחנו מודל EfficientNet-B0 שכבר אומן על 1.28 מיליון תמונות ב-ImageNet, והתאמנו מחדש רק את השכבות האחרונות לסיווג Real/AI. הגישה הזו חוסכת זמן חישוב עצום ומאפשרת ביצועים מעולים גם עם דאטה מוגבל יחסית.



הצינור מעבד תמונת קלט בגודל 224×224 פיקסלים דרך השלד של EfficientNet-B0, מחלץ 1,280 פיצ'רים, ומעביר אותם דרך שכבות Dropout ו-Linear לסיווג בינארי.

למה Transfer Learning? הרשת כבר יודעת לזהות טקסטורות, קצוות ופרטים ויזואליים בסיסיים. אנחנו רק לימדנו אותה מה ההבדל הספציפי בין תמונת AI לתמונה אמיתית — וזה הספיק להגיע ל-98.28% דיוק.

Made with GAMMA

תוצאות המודל — 98.28% Accuracy

98.28%

דיוק כולל

על סט הבדיקה Accuracy

56K

תמונות באימון

מאגר Defactify

66ms

זמן עיבוד

לכל תמונה בודדת

Confusion Matrix

מטריצת הבלבול מציגה את מספר התחזיות הנכונות והשגויות לכל מחלקה. ערכים גבוהים באלכסון הראשי מעידים על סיווג מדויק — המודל מזהה הן תמונות Real והן תמונות AI ברמת דיוק גבוהה, עם שיעור שגיאה נמוך מאוד בשתי הקטגוריות.

GradCAM — מה המודל "ראה"?

מפת החום GradCAM חושפת אילו אזורים בתמונה השפיעו על ההחלטה. צבעים חמים (אדום/כתום) מציינים אזורים בעלי השפעה גבוהה.

- **פנים ועור חלק מדי** → חשד גבוה ל-AI
- **פרטים לא עקביים** → רקע, שיער, ידיים
- **טקסטורות טבעיות** → סימן ל-Real

Made with GAMMA

דמו — נסו בעצמכם

הדמו חי מציג את כל יכולות המערכת בזמן אמת — ללא תלות באינטרנט. מומלץ להכין מראש 2–3 תמונות (אחת AI ואחת אמיתית) כדי להבטיח זרימה חלקה.

01

פתיחת האפליקציה

ניגשים ל-`localhost:3000` בדפדפן — האפליקציה עובדת offline לחלוטין

02

גרירת תמונת AI

המערכת מזהה: **AI** עם ציון ביטחון גבוה + GradCAM שמראה אילו אזורים "הסגירו" את התמונה

03

גרירת תמונה אמיתית

תוצאה: **Real** — השוואה ישירה שמדגימה את יכולת ההבחנה של המודל

04

הצגת Heatmap ↔️ מקור

השוואה בין התמונה המקורית למפת החום — מסביר *למה* המודל הגיע למסקנתו

05

בונוס: ניתוח סרטון

העלאת סרטון קצר — המערכת מנתחת פריימים בודדים ומספקת תוצאה כוללת

Made with GAMMA

מסקנות וכיוונים עתידיים

🚀 כיוונים עתידיים

דאטה בקנה מידה גדול

אימון על **1M+ תמונות** לשיפור כללי ויכולת הכללה

גנרטורים חדשים

הרחבה לתמיכה ב-**Sora,** **Gemini Imagen** וגנרטורי וידאו עתידיים

ציבורי API

שירות פתוח לעיתונאים, חוקרים ומוסדות תקשורת

קוד פתוח

העלאת הפרויקט ל-GitHub עם תיעוד מלא ודוגמאות לשימוש

✅ מה למדנו

EfficientNet-BO — האיזון הנכון

מאזן אופטימלי בין דיוק גבוה למהירות עיבוד בזמן אמת

Transfer Learning יעיל

גם עם דאטה מוגבל, אימון מחדש על מודל קיים מניב תוצאות מצוינות

מוסיף אמון GradCAM

שקיפות בהחלטות המודל הופכת אותו לכלי אמין ונגיש למשתמשים

תודה על ההקשבה!.

Made with GAMMA