

פרויקט 'למידה עמוקה' 2027: סיווג תמונות של עצמים כדוריים

דומים

הבהרה: במקרה של סתירה בין מה שנאמר בכיתה לבין מה שנכתב במסמך זה, הולכים לפי מה שנאמר בכיתה (כפי שהוקלט).

מטרה

המטרה היא לבנות ולהעריך מערכת המסוגלת לסווג עצמים כדוריים מסוגים שונים (כדור, אבטיח, תפוז וכו'), כולל זוגות (או שלישיות וכו') של עצמים שנראים דומים מאוד.

לפרויקט שלושה חלקים:

- נתונים** (אישי): צילום והעלאה של תמונות למאגר המשותף של הקורס.
- קוד** (קבוצתי): השוואת מודלים מוקפאים ו-embeddings (חלקים A ו-B), והגנה עם מצגת.
- איתור שגיאות במאגר** (קבוצתי, רשות): מציאת בעיות במאגר שפורסם.

במהלך העבודה תתנסו ב:

- איסוף נתונים, תיוג ובקרת איכות, כחלק ממאגר תמונות אחד המשותף לכל הקורס.
- שימוש במודלים מאומנים מראש.
- ניתוח ביצועים, ניתוח שגיאות והפקת מסקנות.

הדגש הוא על הבנה, החלטות מנומקות וניתוח התוצאות, ולא רק על המספר הסופי.

נתונים: המאגר המשותף

כל הקורס בונה יחד **מאגר תמונות אחד**, שמשמש את כולם בחלק הקוד. כך מתקבל מאגר גדול ומגוון יותר ממה שקבוצה אחת יכולה לאסוף, וכל הקבוצות מוערכות על אותם נתונים ובאותה חלוקה.

בחירת קטגוריות

- כל סטודנט מציע 3 קטגוריות של עצמים כדוריים, כולל לפחות **זוג "מבלבל"** אחד (למשל תפוז מול כדורסל, אבטיח מול כדור באולינג, כדור טניס מול תפוח ירוק). השראה לזוג מבלבל: ראו את האיור למטה.
- לאחר הצעה (והצבעה אופציונלית), צוות הקורס מאשר רשימה סופית: עצמים עגולים, זמינים לצילום ומבלבלים.



איור: צ'יוואווה או מאפין אוכמניות? זוג מבלבל קלאסי ([Karen Zack](#), @teenybiscuit). חפשו זוגות כאלה בין העצמים הכדוריים. דוגמאות נוספות: [Puppy or bagel](#).

להשוואה, המאגר [Muffin vs chihuahua](#) ב-Kaggle נאסף אוטומטית (scraping) מ-Google Images. אצלנו כל תמונה מצולמת ידנית על ידי סטודנט, מעצם פיזי אמיתי.

📷 צילום והעלאה (אישי)

לפני שמתחילים

- **אחריות אישית:** כל סטודנט אחראי על התמונות שהוא הציע וקיבלו אישור: עמידה במכסה האישית, עמידה בפרוטוקול הצילום ונכונות התיג. התרומה למאגר נמדדת ומקבלת ציון לכל סטודנט בנפרד.
- **מכסה אישית:** 2-3 קטגוריות, 3-4 עצמים פיזיים שונים עבור כל קטגוריה, עם 5 עד 10 תמונות לכל עצם (כ-50-60 תמונות).

⚠️ "עצמים שונים" פירושו פריטים פיזיים שונים: שני תפוזים שונים, ולא אותו תפוז פעמיים.

- **אותו עצם רק בשליחה אחת:** המאגר מחולק ומתויג לפי עצם פיזי. אם אותו תפוז בדיוק יופיע גם ב-train וגם ב-test, התוצאות של כל הקורס יהיו "אופטימיות מדי".

איך מצלמים עצם אחד

1. עצם מרכזי אחד, שנראה לפחות ב-80%, ותופס בין 20% ל-80% מהתמונה.
2. 5 עד 10 תמונות של אותו עצם: לפחות 3 מרחקים (קרוב, בינוני, רחוק) ולפחות 2 זוויות.
3. שני רקע ותאורה בין העצמים: לפחות 2 רקעים ו/או 2 סוגי תאורה בכל קטגוריה שאתם מצלמים.
4. צלמו במצלמת הטלפון הרגילה (JPEG מקורי): בלי פילטרים, בלי חיתוך ובלי צילומי מסך, ברזולוציה מקורית. באיפון: הגדרות ← מצלמה ← פורמטים ← "התאמה מרבית" (Most Compatible), כדי שהתמונות יישמרו כ-

JPEG ולא כ-HEIC.

5. את המגוון יוצרים בצילום, לא בעריכה: אל תעלו עותקים מסובבים, חתוכים או מעובדים של אותה תמונה. זה יתבצע בשלב האימון בלבד.

טוב 	לא טוב 
עצם אחד ברור, ברקעים שונים	כמה עצמים מקטגוריות שונות באותה תמונה
תאורת יום, תאורה מלאכותית, מעט חשוך	חשוך מדי, עד שאי אפשר לזהות את העצם
קצת מקרים קשים: הסתרה חלקית, רקע עמום	כל התמונות על אותו שולחן ובאותה תאורה
רק צילומים שלכם	תמונות מהאינטרנט, צילומי מסך, צילום של מסך או של תמונה מודפסת
בלי אנשים בתמונה, או ידיים בלבד	פנים של אנשים, לוחיות רישוי, מסמכים אישיים

העלאה ומילוי הטופס

- ההעלאה מתבצעת דרך **טופס Google**, ישירות מהטלפון. **כל שליחה = עצם פיזי אחד** ו-5 עד 10 תמונות שלו.
- **קבוצה:** בחירה מהרשימה.
- **קטגוריה:** בחירה מהרשימה.
- **תיאור העצם:** מלל חופשי באנגלית.
- **אישור:** הצילומים שלי, ואני מסכים/ה לפרסומם ברישיון CC BY-NC 4.0 (שימוש חופשי עם ייחוס, לא למטרות מסחריות).

 **טעיתם?** אל תשלחו שוב. ערכו את השליחה דרך הקישור במייל האישור שקיבלתם מהטופס.

 **פרטיות:** המאגר מתפרסם ומתוכנן להיות זמין באופן חופשי. אל תצלמו מקומות או פרטים שאינכם רוצים לשתף!

בסיס נתונים

- עוזר ההוראה מפרסם ב-Kaggle את **גרסה v1.0** המשותפת לכל הסטודנטים: כולל התמונות ו-metadata.
- במקרה של טעויות, גרסת המאגר תעודכן ותופץ הודעה בהתאם.

קוד (קבוצתי, חובה)

הקוד הוא קבצי notebook או py שמראים, על המאגר המשותף, איך מודלים שונים שאומנו מראש וה-embeddings שלהם פותרים את אותה בעיית סיווג. שני החלקים חובה: חלקים A ו-B פותרים את אותה בעיה בשתי דרכים, על אותם נתונים ובאותה חלוקה ל-folds, כדי שההשוואה ביניהם תהיה הוגנת. ההגשה דרך Moodle (ראו "הגשה והגנה").

כללים

- המודלים המאומנים מראש (backbones) משמשים כמחלצי embeddings במצב **מוקפא** (frozen), ללא fine-tuning של המשקלים שלהם.

- מעל ה-embeddings מאמנים מסווג:
 - מסווג קלאסי, כגון SVM, kNN, Logistic Regression, Random Forest.
 - שכבות סופיות חדשות (שכבה ליניארית או MLP קטן) המאומנות עם cross-entropy.
- עיבוד מקדים כמו PCA, נרמול ו-augmentation בזמן חילוף ה-embeddings מותר ורצוי (ראו סעיף Augmentation).

חלק A: Transfer learning מודל supervised לסיווג תמונות

- חילוף embeddings באמצעות מודל שאומן מראש באופן supervised (למשל על ImageNet), ללא שכבת הסיווג שלו.
- השוואה של **לפחות 2 backbones ולפחות 2 מסווגים** מעל ה-embeddings (למשל ResNet-50 מול EfficientNet; SVM מול ראש ליניארי).
- הצגת מרחב ה-(t-SNE / UMAP) embeddings והסבר: אילו קטגוריות מופרדות היטב, אילו לא, ולמה.
- נימוק הבחירה הסופית, כולל ה-trade-off בין דיוק, זמן ריצה וגודל מודל.

חלק B: מודל contrastive / self-supervised של תמונות

- חילוף embeddings באמצעות מודל שאומן מראש (על תמונות בלבד): מודל contrastive (SimCLR, MoCo v3) או מודל self-supervised מסוג (DINO, DINOv2) self-distillation.
- מעל ה-embeddings מאמנים את אותם סוגי מסווגים כמו בחלק A.
- **Few-shot**: השוואה בין חלק A לחלק B כשלכל קטגוריה יש רק k תמונות אימון (למשל k = 1, 5, 10, 20, כולן), בגרף אחד. איזה ייצוג מנצל טוב יותר מעט תוויות? הדוגמאות נדגמות רק מה-folds של האימון; חזרו על הדגימה כמה פעמים (seeds שונים) ודווחו על הממוצע.
- **Retrieval**: לכמה תמונות שאילתה, הציגו את השכנות הקרובות ביותר במרחב ה-embeddings של כל אחד מהחלקים. מה המודל "חושב" שדומה?
- ניתוח מקרים שבהם שתי הגישות טועות באופן שונה, והסבר אפשרי לכך.

Augmentation (בשני החלקים)

- הפעילו augmentations על תמונות האימון (היפוך, סיבוב, חיתוך ושינוי קנה מידה, שינויי צבע ובהירות, טשטוש), חלצו embeddings של העותקים עם אותו backbone מוקפא, והוסיפו אותם לסט האימון של המסווג.
- ה-augmentation מופעל רק על ה-folds של האימון: לעולם לא על ה-fold של ההערכה, וגם לא על חלוקת ה-validation הפנימית.
- השוואה עם ובלי augmentation בתוצאות ה-5-fold, ובמיוחד ב-few-shot. אילו augmentations עוזרות, אילו מזיקות, ולמה? (למשל: שינויי צבע מול תפוז וכדורסל).

דוגמאות למודלים ו-embeddings

אלו נקודות פתיחה, לא רשימה מחייבת. כל מודל מוקפא שאומן מראש מהסוג המתאים מותר; ציינו את שמו המדויק, הגרסה (tag) והמקור.

חלק	מודל	אימון מקדים	ממד ה-embedding
A	ResNet-50	supervised, ImageNet-1k	2048
A	EfficientNet-B0	supervised, ImageNet-1k	1280
A	ConvNeXt-Tiny	supervised, ImageNet-1k	768
A	ViT-B/16	supervised, ImageNet-21k → 1k	768 (וקטור ה-CLS)
B	DINOv2 ViT-S/14, ViT-B/14	self-supervised (self-distillation), ללא תוויות	768 / 384
B	DINO ViT-S/16	self-supervised (self-distillation), ללא תוויות	384
B	MoCo v3 (ViT-B או ResNet-50)	contrastive, momentum encoder	768 / 2048
B	SimCLR (ResNet-50)	contrastive, NT-Xent	2048

טיפים:

- השתמשו בעיבוד המקדים של כל מודל (גודל קלט, mean/std).
- שמרו את ה-embeddings לדיסק פעם אחת לכל backbone (ולכל augmentation), כדי שהניסויים של מסווג לא יריצו את ה-backbone מחדש.

פרוטוקול ניסוי: 5-fold

- כולם משתמשים באותה חלוקה ל-5 folds מהעמודה fold של metadata.csv. החלוקה היא לפי עצם פיזי (object_id): אותו כדור לעולם לא מופיע גם באימון וגם בהערכה.
- בכל סבב מאמנים על 4 folds ומעריכים על ה-fold החמישי. מדווחים Macro-F1 ו-Accuracy כממוצע $\pm$ סטיית תקן על פני 5 הסבבים.
- בחירת hyper-parameters מתבצעת רק בתוך folds האימון (nested cross-validation, או חלוקת validation פנימית, גם כן לפי object_id). ה-fold של ההערכה לא משפיע על שום בחירה.

ניתוח התוצאות אפשרי

הניתוח הוא הבסיס למצגת בהגנה.

- טבלה של Macro-F1 ו-Accuracy (ממוצע $\pm$ סטיית תקן על פני 5 folds) לכל שילוב של backbone ומסווג, בשני החלקים.
- Confusion matrix וביצועים לכל קטגוריה עבור המודל שנבחר בכל חלק, עם דגש על הזוגות המבלבלים.
- גרף t-SNE או UMAP של מרחב ה-embeddings לפחות ל-backbone אחד מכל חלק.
- איור retrieval: תמונת שאילתה והשכנות הקרובות ביותר שלה, בשני החלקים.
- השוואה עם ובלי augmentation.
- Error analysis: גלריה של הצלחות וכישלונות, עם הסבר מדוע כל כישלון קרה.
- ביצועיות: זמן ריצה (ms לתמונה, על CPU או על GPU), גודל המודל וצריכת הזיכרון.

הרחבות רשות

- זיהוי "אף אחת מהקטגוריות" (open-set): המסווג צריך לענות "לא מוכר" במקום לבחור בכוח באחת הקטגוריות. צלמו בעצמכם כמה עצמים עגולים שאינם באף אחת מהקטגוריות (למשל בצל, גלובוס, כדור דיסקו); תמונות אלו לא מועלות לטופס. דחייה לפי סף על המרחק ל-prototype הקרוב ביותר או על ביטחון המסווג.

- **שילוב (fusion) חלקים A ו-B:** האם שני הייצוגים משלימים זה את זה? שרשרו את ה-embeddings של שני החלקים, או בנו ensemble של שני המסווגים, והשוו לכל חלק לבד בהערכת 5-fold, ובמיוחד בזוגות המבלבלים.

איתור שגיאות במאגר (קבוצתי, רשות, 20% ציון מגן)

גם אחרי ה-peer review והניקוי, המאגר שפורסם מכיל טעויות. השתמשו במודלים שלכם כדי למצוא אותן. תיקונים מאושרים ייכנסו לגרסה הבאה של המאגר, והקבוצה תקבל קרדיט.

שיטות אפשריות

- תמונות שרוב השכנות הקרובות שלהן במרחב ה-embeddings שייכות לקטגוריה אחרת.
- תמונות שמסווג (בהערכת 5-fold, כלומר חיזויי out-of-fold) טועה בהן בביטחון גבוה.
- העיניים שלכם: תמיד הסתכלו על תמונה חשודה לפני שאתם מגישים אותה.

מה מגישים: ראו "הגשה והגנה".

הגשה והגנה

- **Moodle:** הקוד הנלווה (notebook או קבצי py) והמצגת.
- **הגנה:** הקבוצה מציגה את הפרויקט במצגת ועונה על שאלות הבוחנים. במסגרת ההגנה יישאלו גם שאלות הקשורות לחומר ההרצאות.
- **Kaggle:** קבוצות במקומות הראשונים בדירוג יתבקשו לעלות את ה-notebook באתר Kaggle, בתיאום ובהנחיית המרצה.

פרטים כלליים

- **גודל קבוצה:** 1-4 סטודנטים, אותן קבוצות כמו בפרויקט הקורס.
- **שימוש ב-LLM ובכלי agent-AI מותר במפורש** (למשל, ChatGPT, Claude / Claude Code, Gemini, Cursor, GitHub Copilot, העוזר המובנה של Kaggle), לכתובת קוד וטקסט.
- עדיין יש צורך להבין את הדברים המהותיים, בפרט אלה שהופיעו בהרצאות. ההבנה נבדקת בהגנה.

הערכה

חלק	מי	מה נבדק	משקל
נתונים	כל סטודנט	תמונות ועצמים שנשארו בגרסה v1.0, עמידה בפרוטוקול ונכונות התיג	30%
קוד	קבוצה	הגנה עם מצגת (35%) ודירוג תחרותי בין הקבוצות (15%)	70%-50%
איתור שגיאות במאגר	קבוצה, רשות	בעיות אמיתיות שאושרו, דיוק הרשימה ושיטת האיתור	20%, ציון מגן

איתור השגיאות אינו חובה. זהו **ציון מגן**: הוא נספר רק אם הוא מעלה את ציון הקבוצה. קבוצה שלא הגישה אותן, מקבלת את ה-20% לקוד (70% קוד).

הדירוג התחרותי בין הקבוצות נקבע לפי התרשמות הבוחנים בהגנה:

- עומק ההבנה של הפרויקט ושל ההחלטות שהתקבלו בו;
- שליטה בחומר ההרצאות שעליו הפרויקט נשען.

⚠ דגשים

- כל תמונה שגויה שזוהתה על ידי סטודנטים אחרים ו/או צוות הקורס מורידה מציון הנתונים באופן משמעותי ביותר!
- תמונות מהאינטרנט, צילומי מסך או תמונות שצילם מישהו אחר מקבלות 0 על כל רכיב הנתונים.
- תמונות שהועלו אחרי סגירת הטופס אינן נספרות.
- **המאגר מתפרסם בפומבי, עם השם של כל אחד מכם.** גרסה v1.0 מתפרסמת ב-Kaggle ברישיון CC BY-NC 4.0, ושמו של כל סטודנט שתרם תמונות מופיע ברשימת התורמים. התמונות שלכם יהיו זמינות לכולם: הקפידו על איכותן ועל הפרטיות.