למידה ללא דוגמאות וקריאת מחשבות

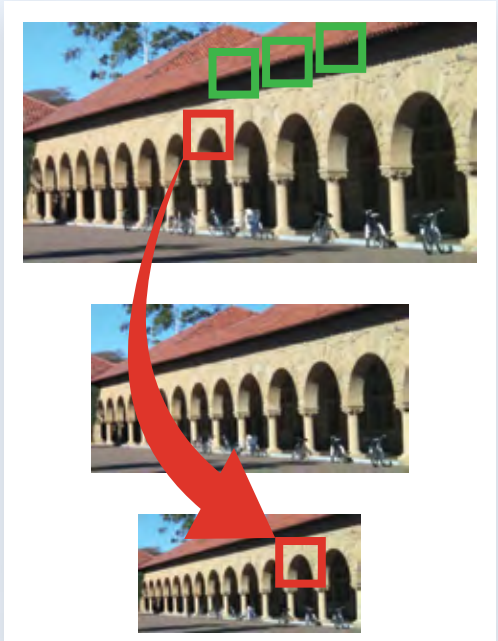
מאת **פרופ' מיכל אירני**

בעשור האחרון הייתה התקדמות עצומה בתחום של אינטליגנציה מלאכותית ולמידת מכונה, ובייחוד בתחום הלמידה העמוקה (deep learning) ורשתות עצביות מלאכותיות (artificial neural networks). מחשבים מסוגלים לבצע כיום משימות שלא העלינו אותן על דעתנו לפני עשור. הם מסוגלים לזהות פרצופים טוב יותר מבני אדם, להסיע מכונית ללא נהג, לנצח את אלוף העולם בשחמט ובגו (Go), לכתוב טקסטים מופלאים באמצעות ChatGPT ועוד. כל הדברים הנפלאים הללו לא היו מתאפשרים לולא קיומו ונגישותו של אוסף **עצום** של נתונים דיגיטליים. למעשה, כדי שמערכות ממוחשבות יצליחו לבצע משימות אלו בהצלחה, נדרשים להן מיליונים רבים של דוגמאות לימוד. למשל, מערכות מתקדמות לזיהוי פרצופים למדו קרוב למיליארד פרצופים מתויגים על פני הזמן, כלומר שיעור נכבד מאוכלוסיית העולם. לא פלא שהן עושות זאת טוב יותר מבני אדם...

אבל נשאלת השאלה מה קורה כשאין מספיק דוגמאות (למשל בתחומים רפואיים למיניהם), או כשהדוגמה החדשה איננה במרחב ההתפלגות של הדוגמאות שלימדנו את המחשב. במצבים כאלה אינטליגנציה מלאכותית ומערכות ממוחשבות נכשלות כישלון מחפיר. לעומתן, בני אדם מסוגלים להכליל ממעט מאוד דוגמאות. למשל, כשמראים לילדים קטנים תמונות מעטות של

פרה (חלקן אפילו מצוירות), הם מסוגלים לזהות פרה כשהם רואים אחת באחו. איננו צריכים ללמד אותם אלפי תמונות של פרות, מכל הזוויות, בכל הגדלים, בקומבינציות שונות של כתמים, כדי שהם ידעו לזהות פרה. הבדל זה הוא רק אחד מני רבים הממחישים את העובדה שאינטליגנציה אנושית ואינטליגנציה מלאכותית שונות מאוד זו מזו.

אחת השאלות הגדולות הפתוחות בתחום האינטליגנציה המלאכותית היא כיצד בכל זאת ניתן ללמד מחשב לבצע היסקים הגיוניים על סמך דוגמאות לימוד מעטות בלבד. מאמרי זה מציג שניים ממגוון ערוצי המחקר שנחקרים בקבוצתי (במכון ויצמן למדע) העוסקים בשאלה זו. ערוץ מחקר אחד בוחן כיצד ניתן ללמד מחשב לבצע משימות מורכבות על סמך דוגמת לימוד בודדת



איור 1. חזרתיות של "פאצ'ים" בתוך תמונה (within & across scales). פאצ'ים נוטים לחזור על עצמם פעמים רבות בתוך תמונה בודדת – הן בתוך התמונה בגודלה המקורי (within scale – מודגם בפאץ' ירוק) והן בגרסאות מוקטנות שלה (across scales – מודגם בפאץ' אדום). אף שתמונה ספציפית זו מאופיינת בחזרתיות ברורה לעין, ומודגמת על פאצ'ים גדולים, מתברר שאם בוחנים פאצ'ים קטנטנים, חזרתיות זו תקפה כמעט לכל פאץ' קטן כמעט בכל תמונה.

(תמונה אחת בלבד). הערוץ השני עוסק ב"קריאת מחשבות" (פיענוח מידע ויזואלי שאדם ראה, ישירות מתוך הקריאות המוחיות שלו), ואנו מראים כיצד ניתן ללמד מחשב לבצע זאת על אף מיעוט דוגמאות הלימוד.

כיצד ניתן ללמוד מתמונה בודדת?

בתמונה בודדת יש חזרתיות (יתירות) רבה של מידע. פיסות קטנות של התמונה חוזרות על עצמן פעמים רבות – הן בסקאלה המקורית של התמונה והן בסקאלות אחרות (גדלים שונים) של התמונה. למה הכוונה? קרוב מאוד לוודאי שלכל פיסת מידע קטנה שמופיעה בתמונה (לשם פשטות נקרא לה "פאץ'" – **image patch**) יופיעו עוד הרבה פאצ'ים הזהים לה במקומות אחרים באותה התמונה. איור 1 מדגים חזרתיות כזו של פאצ'ים גדולים (המודגשים בצבע ירוק), אך בפועל הכוונה לפאצ'ים הרבה יותר קטנים. חזרתיות כזו מתרחשת גם בין סקאלות שונות של התמונה (**across scales**), כפי שמודגם באיור 1 באמצעות פאצ'ים המסומנים בצבע אדום. כמעט לכל פאץ' בתמונה המקורית ניתן למצוא פאצ'ים כמעט זהים לו בגרסאות מוקטנות של התמונה (בלי להקטין את הפאץ'), מה שאומר שהפאץ' המקורי מופיע בכל מיני גדלים בתמונה המקורית. התמונה שמוצגת באיור 1 היא בעלת חזרתיות מובנית וברורה לעין, וניתן לשאול: **האם חזרתיות כזו קיימת בכל תמונה באשר היא?** מתברר שאם בוחנים פאצ'ים קטנים מספיק (5x5 או 7x7 פיקסלים), התשובה היא "כן". במאמר מ־2009 הראינו שחזרתיות כזו (גם **within** וגם **across scales**) תקפה כמעט לכל פאץ' קטן כמעט בכל תמונה טבעית. זוהי תכונה ממש מופלאה של אינפורמציה ויזואלית(!). חזרתיות חזקה זו שימשה אותנו, כמו גם חוקרים אחרים, בפתרון בעיות רבות בראייה ממוחשבת. להלן אתאר בקצרה ארבעה שימושים מתוך מגוון השימושים הרבים שמצאנו לתכונה זו: ◄

איור 2. השלמת מידע חסר בתמונה (באמצעות מקסום החזרתיות הפנימית של פאצ'ים)

1) **השלמת מידע חסר**: נניח שרוצים להשמיט חלק מתמונה קיימת ולהשלים את החסר במקום אחר בה, למשל להסיר את קופץ הבנג'י מתוך התמונה שבאיור 2 ולהשלים את מה שהיה מאחוריו. מובן שאיננו יודעים בפועל מה היה מאחוריו, ויש פתרונות רבים מספור למה שיכול היה להיות שם. לפיכך אנו רוצים לספק את ההשלמה **הסבירה ביותר**. אנו הגדרנו את ההשלמה "הסבירה ביותר" כך: השלמה שתמקסם את החזרתיות של הפאצ'ים בתוך התמונה (במגוון סקאלות של התמונה), ופיתחנו אלגוריתם שזו משימתו. באיור 2 ניתן לראות דוגמה לתוצאה של הפעלת האלגוריתם. זו ההשלמה שסיפקה חזרתיות של הפאצ'ים בתמונה באמצעות השלמת החלק החסר (כמו "פאזל") בחתיכות קטנות שנלקחו ממקומות אחרים בתמונה. למעשה, האפליקציה Content-Aware Fill של Adobe Photoshop מבוססת על האלגוריתם שלנו (מכון ויצמן למדע מכר ל־Adobe זכויות על הפטנט). את השלמת החסר אפשר לבצע כמובן לא רק בתמונות אלא גם בסרטים. במקרה זה, פיסות המידע החסרות (חתיכות הפאזל) הן חתיכות זמן מרחב קטנות (space-time patches) הנלקחות מחלקים אחרים של הווידאו. דוגמאות של השלמות וידאו אפשר למצוא באתר האקדמיה, בהרצאת הבכורה שלי עם הצטרפותי לאקדמיה.

2) **סופר־רזולוציה**: הרזולוציה של תמונה תלויה בצפיפות הפיקסלים שבתוכה (למשל כמה מגה־פיקסל יש בתמונה). כשמגדילים את הרזולוציה של תמונה, מגדילים את צפיפות/מספר הפיקסלים שבה, וזה מתבטא בהגדלת התמונה עצמה. אולם איננו רוצים להגדיל את התמונה רק באמצעות מתיחתה, שלצורך זה משמשת ה"אינטרפולציה" (שיטה למתיחת התמונה לתמונה גדולה יותר, אך אינה

תמונת קלט (ברזולוציה נמוכה)

הגדלה באמצעות סופר־רזולוציה

הגדלה באמצעות אינטרפולציה (מתיחה)

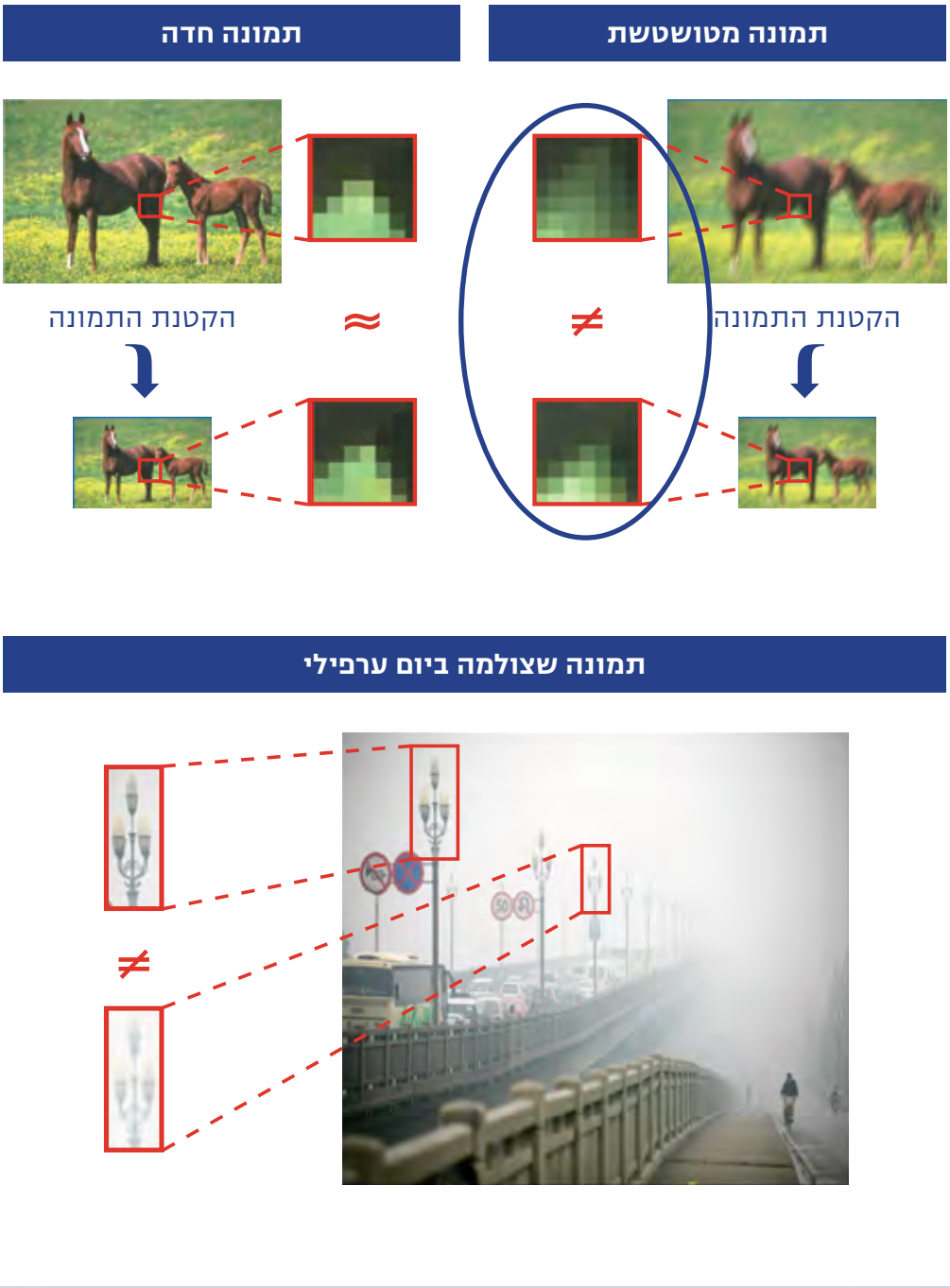
איור 3. סופר־רזולוציה (מתמונה אחת)

מוסיפה מידע חדש). להוספת מידע חדש בהגדלת התמונה משמשת שיטה המכונה "סופר־רזולוציה", שבאמצעותה נוספים פרטים חדשים **הקטנים מגודל של פיקסל בתמונה המקורית** (beyond the Nyquist limit); פרטים בגודל כזה אי אפשר לשחזר או לראות במתיחה (אינטרפולציה). כזכור, כל מה שיש היא תמונה בודדת, בעלת רזולוציה מוגבלת בלבד, ואנו מניחים שאין דוגמאות נוספות, אך מתברר שאנחנו יכולים לשחזר מתמונה בודדת זו תמונה חדשה בעלת רזולוציה גבוהה יותר, בשימוש בחזרתיות הפנימית הקיימת בתמונה עצמה. כזכור, כל פאץ' חוזר הרבה פעמים בתוך התמונה באותה הסקאלה (פאצ'ים ירוקים באיור 1) באופן בלתי נמנע הפאץ' חוזר בהזזות **תת־פיקסליות**. לפיכך כל החזרות הללו יחד מספקות הרבה יותר דגימות של אותו הפאץ'. אם נשלב את כל הדגימות הללו ביחד, הרי שנקבל בפועל רזולוציה גבוהה יותר של הפאץ'. יתרה מזו – כזכור, פאץ' מופיע גם בכל מיני גדלים בתמונה (בסקאלות שונות, כגון הפאץ' האדום באיור 1). כלומר, למעשה יש בתוך התמונה עצמה דוגמאות לפאץ' זה ברזולוציות שונות; דוגמאות לאיך הפאץ' אמור להיראות אם נגדיל את התמונה. כזכור, כשהפאצ'ים מספיק קטנים (5x5 או 7x7 פיקסלים) חזרתיות זו תקפה לכל פאץ' בכל תמונה טבעית. משמעות הדבר היא שניתן לבצע סופר־רזולוציה לכל תמונה שהיא, גם כשנתונה רק התמונה עצמה, ללא דוגמאות נוספות(!). איור 3 (ראו לעיל בעמ' 11) מדגים תוצאה של סופר־רזולוציה מתמונה בודדת.

ב־2009 הצגנו לראשונה אלגוריתם לסופר־רזולוציה מתמונה בודדת (המבוסס על החזרתיות הפנימית) בשיטות של אופטימיזציה קלאסית. ב־2018 הרחבנו רעיון זה ללמידה עמוקה. הראינו שהחזרתיות הפנימית בתוך כל תמונה מאפשרת **לאמן רשת נֵירונים על תמונה בודדת**. קראנו לזה deep internal learning. בתחילה הדגמנו יכולת זו בבעיית הסופר־רזולוציה, ובהמשך הרחבנו את הגישה והדגמנו את יישומה במגוון רחב של אפליקציות שונות.

עד כה ראינו שכמעט לכל פאץ' בתמונה המקורית יש פאצ'ים זהים לו בגרסאות מוקטנות של התמונה. אך מתברר שהחזרתיות הפנימית המופלאה הזו קיימת רק בתנאים "אידיאליים", כלומר רק בתמונות נקיות וחדות. לעומת זאת בתמונות "מקולקלות" (למשל תמונות מטושטשות או תמונות שצולמו בתנאי ערפל וכו') פאצ'ים שהיו זהים אילו צולמה התמונה בתנאים אידיאליים, אינם זהים זה לזה. הקלקול (של העדשה במקרה של טשטוש, או של העולם הפיזיקלי במקרה של ערפל) גורם לירידה ניכרת בדמיון הפנימי של הפאצ'ים בתמונה (איור 4). מתברר שסטייה זו מחזרתיות מושלמת **מקודדת** בתוכה מידע על הקלקול שאחראי לתוצאה שבתמונה. אנו פיתחנו תאוריה הנקראת blind optics, שמאפשרת לחשב מתמונה בודדת (ללא שום דוגמאות נוספות) את הקלקול שלפיו נוצרה התמונה. הקלקול מחושב לפי מידת **הסטייה** מהחזרתיות המושלמת של הפאצ'ים (within and across scales).

3) **הסרת טשטוש מתמונה**: כאשר מקטינים תמונה מטושטשת, היא נעשית חדה יותר. לכן במקרה של טשטוש אנו מחפשים פונקציית טשטוש, ואם נסיר באמצעותה טשטוש זה מהתמונה, נמקסם את הדמיון הפנימי של הפאצ'ים בין סקאלות שונות (גדלים שונים) של התמונה שנוצרה.



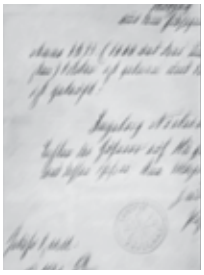
איור 4. הסטייה מחזרתיות מושלמת מקודדת את הקלקול.
פאצ'ים בתמונה איכותית (למשל תמונה חדה) כאמור חוזרים על עצמם בסקאלות שונות של התמונה, אולם חזרתיות זו נשברת בתמונה "מקולקלת" (מטושטשת, רועשת, מעורפלת וכו'). סטייה זו מדמיון מושלם של פאצ'ים מקודדת בתוכה את סוג הקלקול שבתמונה ומאפשרת את חישוב הקלקול עצמו לצורך תיקון התמונה.

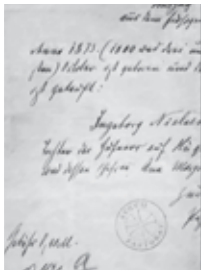
(א) הסרת טשטוש מתמונות

התמונה המטושטשת (הקלט)

פונקציית הטשטוש שחושבה מהתמונה

התמונה החדה ששוחזרה (הפלט)

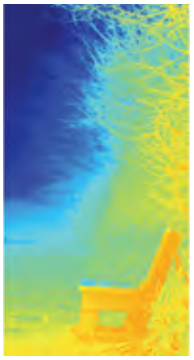




(ב) הסרת ערפל מתמונות

תמונה שצולמה ביום ערפילי (הקלט)

שחזור מפת עומק מהתמונה



הסרת הערפל מהתמונה (הפלט)

איור 5. שימוש בסטייה מחזרתיות מושלמת כדי לחשב את הקלקול

א) הסטייה מחזרתיות מושלמת מאפשרת לחשב את פונקציית הטשטוש (מסומנת במסגרת אדומה). לאחר שחזור פונקציית הטשטוש ניתן להסירה מהתמונה (באמצעות פעולה הנקראת "דקונבולוציה") לשם קבלת תמונה חדה.

ב) הסטייה מהחזרתיות מאפשרת לחשב הן את מפת העומק של הסצנה (כחול = רחוק; כתום = קרוב) והן את הפרמטרים של הערפל. בהינתן העומק והפרמטרים של הערפל ששוחזרו מהתמונה, ניתן להשתמש בנוסחה פיזיקלית של ערפל כדי להסיר את הערפל מהתמונה ולקבל תמונה שכאילו צולמה ביום שהראות בו הייתה טובה.

דוגמה לפונקציית הטשטוש שחושבה מתמונה מטושטשת ולתמונה שנוצרה לאחר הסרת הטשטוש ניתן למצוא באיור 5א.

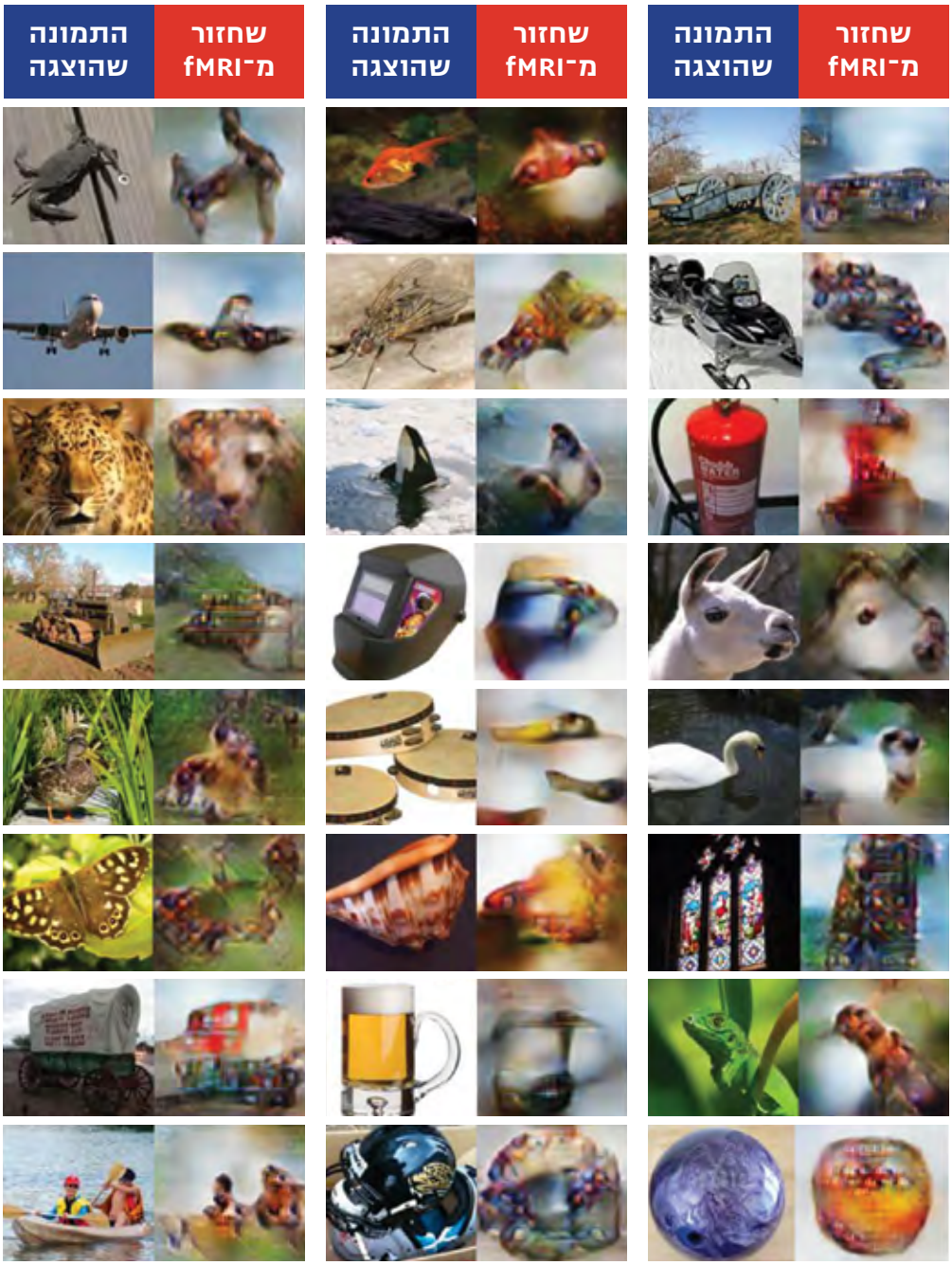
4) **הסרת ערפל מתמונה**: ביום שהראות בו טובה, פרטים דומים בעומקים שונים בסצנה אמורים להיראות אותו דבר. פרטים דומים בעומקים שונים נראים שונה (למשל, הפנסים באיור 4, או הענפים באיור 5ב). סטייה זו מדמיון מושלם מאפשרת לחשב את העומק היחסי שבין שני פאצ'ים אלה. למעשה, הסטייה מהחזרתיות של כל הפאצ'ים בתמונת ערפל בודדת מאפשרת לחשב הן את מפת העומק של הסצנה כולה (איור 5ב) והן את הפרמטרים של הערפל. בהינתן העומק והפרמטרים של הערפל ששוחזרו מהתמונה, ניתן להשתמש בנוסחה פיזיקלית של ערפל כדי להסיר את הערפל מהתמונה ולקבל תמונה שכאילו צולמה ביום שהראות בו הייתה טובה (איור 5ב). ניתן לראות שבתמונה ששוחזרה ללא ערפל הדמיון הפנימי של הפאצ'ים רב במידה ניכרת מהדמיון הפנימי שבתמונת הערפל המקורית.

"קריאת מחשבות"

ערוץ מחקר נוסף במעבדתי עוסק ב"קריאת מחשבות", כלומר פיענוח מידע ויזואלי שאדם ראה, ישירות מתוך הקריאות המוחיות שלו. גם במקרה זה נתון לנו רק מספר מצומצם של דוגמאות לימוד (כפי שיוסבר בהמשך). הסיטואציה היא כלהלן: אדם שוכב בתוך מכונת fMRI שמקליטה את הפעילות המוחית שלו בעודו צופה בתמונות או בסרטי וידאו. חשוב להדגיש שלא הוחדר דבר למוח; אלה רק צילומים. כמו כן יש לציין שמכונת fMRI איננה מודדת את הפעילות העצבית במוח ישירות אלא רק זרימת דם מחומצן לאזורים שהם פעילים במוח יותר מאחרים, ולכן זהו מידע עקיף מאוד ו"רועש" מאוד. במחקרנו הנוכחי אנו בוחנים את המדידות שהוקלטו באזור הראייתי של המוח (visual cortex) ומנסים לשחזר מהן את התמונה/הסרט שהאדם ראה בזמן שהן הוקלטו. איור 6 (ראו להלן בעמ' 16) מדגים שחזורים מתוך קריאות מוחיות: בכל זוג תמונות באיור התמונה השמאלית היא זו שהאדם צפה בה, והימנית היא זו ששחזרנו מהפעילות המוחית שלו שהוקלטה במכונת fMRI בזמן הצפייה.

חשוב להדגיש שאין מה להיבהל – מחשבות אינן "מתרוצצות באוויר". יש צורך בשיתוף פעולה אקטיבי ומאומץ של האדם השוכב במכונת ה־fMRI. עם זאת, בשל יכולותיה של "קריאת מחשבות" שכזו אפשר שהיא תשמש בכמה תחומים בטווח הרחוק יותר, למשל: בממשקי אדם–מכונה, לעזור בתקשורת עם אנשים שאינם יכולים לדבר ולתקשר באופן פיזי אחר (למשל חולי ALS או אנשים בתרדמת), "לצפות" בחלומות של אנשים וכו'. מחקרנו עדיין לא שם, אך אנו עושים צעדים ראשונים בכיוונים אלה.

האלגוריתם שפיתחנו לצורך פיענוח התמונות מקריאות מוחיות מבוסס על רשת נירונים, המקבלת כקלט קריאת fMRI (פעילות מוחית) ומוציאה כפלט את התמונה שיצרה פעילות מוחית זו. לשם הפַּשטוּת נקרא לרשת זו decoder. אילו היו לנו מיליוני דוגמאות של תמונות בליווי הפעילויות המוחיות המוקלטות שלהן לא הייתה בעיה לאמן את ה־decoder על דוגמאות רבות מאוד ולהגיע לביצועי פיענוח טובים מקריאות מוחיות חדשות שטרם נצפו. הבעיה היא שאין הרבה דוגמאות... אדם יכול לשהות בתוך מכונת fMRI לפרק זמן מוגבל בלבד, מה שמגביל את מספר דוגמאות הלימוד שאנו יכולים לאסוף. מספר הדוגמאות שנאספו והיו נגישות לנו בעת אימון ה־decoder היו כאלף תמונות בלבד (בליווי הקריאות המוחיות שלהן) – ◄



איור 6. "קריאת מחשבות" (שחזור תמונות מפעילות מוחית)
בכל זוג תמונות באיור התמונה השמאלית היא זו שהאדם צפה בה, והימנית היא זו ששחזרנו מהפעילות המוחית שלו שהוקלטה במכונת fMRI בזמן הצפייה.

מספר דוגמאות מצומצם מאוד, שבוודאי אינו מכסה את מרחב כל התמונות האפשריות שאדם עשוי לצפות בהן בעתיד. לפיכך רשת שאומנה על אלף דוגמאות בלבד לא תדע להכליל על דוגמאות חדשות לכשתפגוש בהן.

כדי לפתור בעיה זו של מיעוט דוגמאות לימוד אפשרנו ל־decoder "ללמד את עצמו", כפי שתינוק מלמד את עצמו בניסוי וטעייה (self supervised training). לשם כך נוסף על רשת ה־decoder, שאמורה ללמוד **לפענח** תמונות מקריאות fMRI, אימנו עוד רשת – encoder – שמטרתה לעשות את הפעולה ההפוכה – כלומר ללמוד **לקודד** תמונות ל־fMRI. שילוב של שתי רשתות אלו, ההפוכות זו לזו, מאפשר להפעיל "לימוד עצמי" של המערכת, כלהלן: אם ניקח תמונה כלשהי שמעולם לא נצפתה בתוך מכונת fMRI (ולכן אין עבורה קריאה מוחית) ונזין אותה ל־encoder שלנו, הוא יקודד אותה ל־fMRI כלשהו (לאו דווקא נכון). כעת, אם נפעיל על ה־fMRI שנוצר את ה־decoder שלנו, הוא מן הסתם ייצר תמונה "זיבלית" (כי לא התאמן על מספר גדול דיו של דוגמאות). אולם אילו היו שתי הרשתות הללו מאומנות כראוי, הרי שההלוך־ושוב הזה (מתמונה ל־fMRI, ובחזרה מ־fMRI לתמונה) היה אמור תאורטית להחזיר אותנו לתמונה שהתחלנו בה. לפיכך אנחנו יכולים לאמן את שתי הרשתות הללו (ה־encoder וה־decoder) יחד, ואז המעבר המעגלי הזה הלוך־ושוב דרך שתיהן יחויב להשיב **כל תמונה** לעצמה. שימו לב שאת התהליך הזה של ההלוך־ושוב המעגלי אנו יכולים להפעיל על **כל תמונה שהיא**, כולל תמונות שאין עבורן שום מדידת fMRI. משמעות הדבר שאנו יכולים לאמן את צמד הרשתות הללו על "**אין־סוף**" תמונות, ואיננו מוגבלים רק לאלף דוגמאות אימון(!). אימון כזה על אין־ספור תמונות טבעיות (נוסף על אלף דוגמאות האימון עם קריאות ה־fMRI) מאפשר ל־decoder שלנו להתאים את עצמו למרחב העצום של תמונות טבעיות אף שמעולם לא נצפו במכונת fMRI.

סיכום

במאמר זה אפשרתי הצצה ממעוף הציפור לשניים ממגוון ערוצי המחקר בקבוצתי. ערוצים אלה חוקרים כיצד ניתן לאמן מערכות ממוחשבות ולאפשר להן לבצע משימות מורכבות גם במצבים שבהם יש מעט מאוד דוגמאות אימון. יתרה מזאת, ראינו כי אפילו במצבים שבהם אין שום דוגמה מוקדמת, בכל זאת ניתן לפתור בעיות מורכבות באמצעות אימון הרשת על תמונה בודדת – תמונת הקלט עצמה (תוך כדי ניצול החזרתיות הפנימית שבתמונה). ■