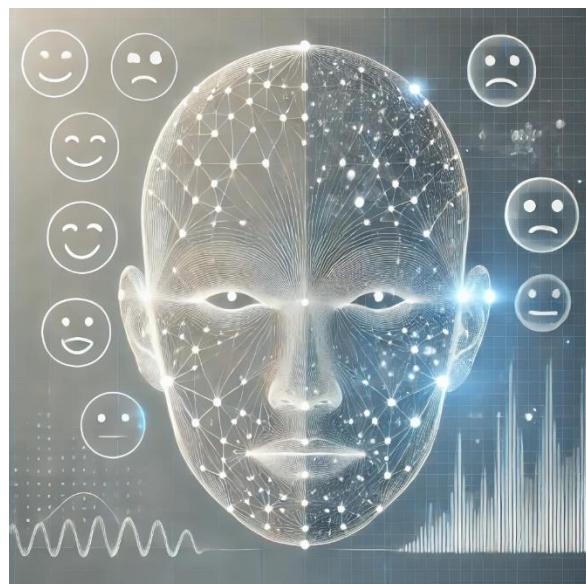


# **סמינר בביונה מלאכותית**

**זיהוי רגשות באמצעות למידה عمוקה**



**עידו עמר, 204682769**

**בנהנויות ד"ר מיה הרמן**

**תוכן עניינים:**

3.....	פתח דבר .....
4.....	פרק 1 - מבוא .....
4.....	רקע כללי על למידה عمוקה .....
4.....	למידה عمוקה בזיהוי רגשות .....
5.....	חשיבות יכולת זיהוי רגשות .....
5.....	הבעיה המחקרית .....
5.....	מטרות ושאלות המחקר .....
7.....	פרק 2 - שימושים אפשריים בזיהוי רגשות .....
9.....	פרק 3 - מודלים וטכניקות בלמידה عمוקה לזיהוי רגשות .....
9.....	א. הכנת הנתונים ועיבוד מקדים .....
9.....	מארגני נתונים .....
12.....	זיהוי הפנים - כלים וטכניקות .....
13.....	תיקון תМОנות וניקוי רשעים .....
15.....	ב. זיהוי רגשות .....
16.....	ארכיטקטורה של CNN .....
25.....	שימוש CNN לזיהוי רגשות .....
27.....	סקירת מודלים .....
33.....	ג. אופטימיזציה .....
35.....	פרק 5 - מסקנות ודיון .....
37.....	פרק 6 - סיכום אישי .....
38.....	פרק 7 - רישימת מקורות .....

### **פתח דבר:**

תחום הבינה המלאכותית עבר התקדמות מהירה ומשמעותית של מחקר ויישום רבים, ובפרט על זיהוי רגשות באמצעות מכונה. זיהוי רגשות פנים משמש כויס בוגון רחב של יישומים, החל מערכות לשיפור חוותית משתמש ועד ניטור מצלבים רגשיים בהקשרים רפואיים וחינוכיים. העבודה הנווכחית מתמקדת בחקר ובשימוש של רשותות ניירונים קוונטוציוניות (CNN) למשימת זיהוי רגשות, תוך בחינת ארכיטקטורות שונות, שיטות עיבוד נתונים ואופטימיזציה של היפר-פרמטרים.

כתיבת העבודה סיפקה לי הזדמנות להעמק בתחום הלמידה העמוקה ולבחון את האתגרים הכרוכים בישום טכנולוגיות מתקדמות אלה. ההתמודדות עם בחירת מודלים מתאימים, הכנת נתונים, ואופטימיזציה חשפה את המורכבות הטכנולוגיות והמתודולוגיות של התחום. מעבר להיבט המחקרי, העבודה העניקה לי כלים פרקטיים, אשר אני עתיד לישם בסדנה המדעית הנתונים שבה אני לוקח חלק, שבה בחרתי ממש מודל לזיהוי רגשות.

העבודה מציגה את המשען המחקרי שuberתני, החל מההיכרות עם התחום, דרך סקירת המודלים והארQUITECTURES השונות, ועד לבחינת השפעתן של שיטות אופטימיזציה על הביצועים.

פרק 1 - מבוא

רקע כללי על למידה عمוקה

למידה عمוקה (Deep Learning) היא תת-תחום של למידה מכונה המבוססת על רשתות נוירונים מלאכותיות עמוקות. טכנולוגיה זו מאפשרת למערכות ממוחשבות ללמידה וולזהות ובניות מורכבות מותוך כמיות גודלות של נתונים, ללא צורך בהנחיות מפורשות מאות משתמשים.

בשנים האחרונות, למידה عمוקה חוללה מהפכה במגוון רחב של תחומים, כמו ראייה ממוחשבת, עיבוד שפה טבעית (NLP), ורוביוטיקה.

רשתות נוירונים קונבולוציוניות (CNN - Convolutional Neural Networks) הן טכנולוגיה מרכזית  
בלימידה عمוקה, במיוחד בתחום תמונה. הן מאפשרות זיהוי מדויק של אובייקטים, טקסטורות ודפוסים  
חוותיים על ידי שימוש במידע בשכבות היררכיות. טכנולוגיות מתקדמות אחרות, כמו רשתות חוזרות  
תובנות דינמיות על בסיס נתונים משתנים לאורך זמן.  
(RNN - Recurrent Neural Networks)

בזכות השיפור בעוצמת המחשב והזミニות של מערכיו נטוניים גדולים, למידה عمוקה היפה לכלי עוצמתי שיכול להתמודד עם בעיות מורכבות. שילוב של למידה عمוקה עם הנדסת תוכנות וטכניקות אופטימיזציה מתקדמות.אפשרויות לשפר את הביצועים ואת דיקט המודלים.

למידה عمוקה בזיהוי רגשות

אחת היכולות של למידה عمוקה היא יכולת לזיהות רגשות מותוך מגוון מקורות מידע, כמו טקסט, קול, ותמונות. זיהוי רגשות (Emotion Recognition) הוא תחום מחקר מתפתח שמטרתו להבין ולנתה את המצב הרגשי של אנשים. רגשות הם מרכיב מרכזי בתקשורת האנושית, וזיהוי מדויק שלהם מאפשר שיפור באינטראקציה בין אדם למוכנה ובין אדם לאדם.

זיהוי רשות חזותי, המבוסס על ניתוח הבעות פנים, הפך לתחום פופולרי בזכות היכולת של רשות CNN לנתח תמונות ולזהות תכונות מורכבות בהבעות פנים. רשות כגון שמחה, עצב, פחד וכעס מתבטאים בהבעות פנים שניתנו לזהותן באמצעות מודלים מתקדמים.

מעבר לזהויו רגשות בסיסיים, למידה עמוקה מאפשרת ניתוח רגשות מורכבים יותר באמצעות שילוב מידע חזותי עם מקורות נתונים נוספים, כגון קול (לדוגמא, אינטונציה בתוך הדיבור) או טקסט (לדוגמא, ניתוח המלל). מודלים רב-מודולריים (Multimodal) אלו מספקים תובנות עשירות יותר על מצבים ורגשיים מורכבים.

## **חשיבות יכולת זיהוי רגשות**

זיהוי ורגשות הוא תחום מחקר מתחם המשלב טכנולוגיות מתקדמות של למידה عمוקה כדי לאפשר הבנה עמוקה יותר של רגשות אנושיים. הרגש הוא אחד המאפיינים הבולטים של התקשורות האנושית, והוא משפייע באופן משמעותי על התנהגות, קבלת החלטות וឥינטראקטיביות חברתיות.

היכולת לזהות רגשות בזמן אמת פותחת דלתות ליישומים חדשניים בתחוםים מגוונים, בהם חינוך, רפואי, שיווק, ואבטחה.

בתחום החינוך - במערכות למידה מקוונת, שבון חסנה נוכחות פיזית של מורים ותלמידים, זיהוי רגשות בזמן אמת יכול לספק תובנות חשובות על מעורבותם והבנתם של התלמידים. לדוגמה, מערכת שמזהה תלמידים אשר אינם מעורבים יכולה לעזור למורים להתאים את שיטות ההוראה ולהגבר את העניין של התלמידים בחומר הנלמד.

תחום הבריאות והרפואה - רגשות שימושיים כלים חשובים לאבחן מצבים נפשיים. מערכות שיכלות לזהות רגשות בצורה אוטומטית עשוות לסייע באבחון מוקדם של דיכאון, חרדה, או מצבים נפשיים אחרים. פתרונות מבוססי זיהוי רגשות יכולים לשפר את חיويות המטופל ולהקל על אנשי מקצוע בתחום הבריאות באבחון ומתן טיפול מותאם אישית.

תחום השיווק והבידור - במרחב העסק, זיהוי רגשות מאפשר לחברות להבין את התגובה הרגשות של לקוחות למוצרים ושירותים. כך ניתן להתאים חיויות משתמש וליצור פתרונות אישיים יותר. בתעשייה הבידור, מערכות לזיהוי רגשות מאפשרות ייצור תוכן שמותאם למצבים ונסיבות של המשתמשים.

באבטחה ובטיחות - במערכות אבטחה מתקדמות, זיהוי רגשות יכול לשמש לזיהוי רגשות חריגים או מצבים רגשיים קיצוניים, כמו פחד או כאס, בזמן אמת. יישומים כאלה עשויים לסייע במניעת אירועים מסוכנים למרחבים ציבוריים או פרטיים.

## **הבעיה המחקרית**

למרות ההתקדמות המרשימה בלמידה عمוקה, קיימות מספר מגבלות ואתגרים בתחום זיהוי רגשות בזמן אמת:

- דיק נזוק בתנאים מורכבים - מודלים רבים מתקשים לזהות רגשות בתנאים של תאוריה משתנה, רעים סביבתיים או זוויות פנים שונות.
- זמן עיבוד - כדי לזהות רגשות בזמן אמת, יש צורך במערכות מהירות וחסכנות מבחן חישוב.
- הטויות בתנאים - מודלים רבים מתבססים על מרכי נתונים שאינם מוגנים מספיק מבחינה תרבותית או מדנית, מה שעולול להוביל להטויות בתוצאות.
- אתגרים טכנולוגיים ואתיים - שימוש במערכות זיהוי רגשות מעלה שאלות של פרטיות ואתיקה, במיוחד בסביבות רגשות כמו חינוך ובריאות.

## **מטרות ושאלות המחקר**

מטרות המחקר:

- לחקר טכנולוגיות מתקדמות לזיהוי רגשות בזמן אמת - הבנת הבעיות והמגבלות של שיטות למידה عمוקה, כגון CNN ו-ResNet.
- שיפור דיק ומהירות - סקירת מודלים המשלבים טכניקות עיבוד מתקדמות, כמו אופטימיזציה מבוססת Walrus.

- בדיקת יישומים מעשיים - בחינת השימושים האפשריים של מערכות זיהוי וגישה בתחוםים שונים, כמו חינוך, רפואי, ושיווק.
- הבנת אתגרים אתיים - ניתוח השפעתן של מערכות אלה על פרטיות המשתמשים והדרכים לצמצם סיכוןים.

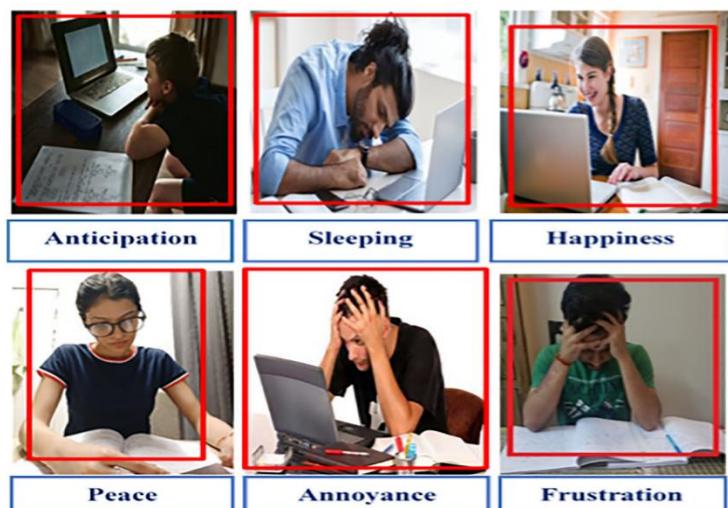
**שאלות המחקר :**

1. כיצד ניתן לשפר את הדיווק וה מהירות של מודלים לזיהוי וגישה בזמן אמיתי?
2. איזה מודל למידה عمוקה, מבין המודלים הניסקיים, הוא האפקטיבי ביותר בזיהוי וגישה חזותי?
3. כיצד ניתן לשלב נתוני וידאו וקול לייצרת ניתוח גישה מדויק ורב-מודאלי?
4. מהן המוגבלות האתניות והטכנולוגיות של מערכות לזיהוי וגישה, וכיים ניתן להתמודד איתן?

## **פרק 2 - שימושים אפשריים בזיהוי רגשות**

למידה מקוונת הפכה לחלק בלתי נפרד מערכות החינוך המודרניות, אך היא מציבה אתגר משמעותי בזיהוי מידת המעורבות וההבנה של התלמידים בהיעדר אינטראקציה פיזית ישירה עם המורה. במצבים של הוראה פרונטלית, מורים יכולים לזהות בקלות סימנים המעידים על תשוכן, חוסר ריכוז או חוסר עניין, אך במרחבים וירטואליים היכולת זו מוגבלת. כדי להתמודד עם בעיה זו, פותחו מערכות מבוססות למדת מכונה שמאפשרות מדידה אובייקטיבית של מצבם הרגשי של התלמידים באמצעות ניתוח בזמן אמת הבאות פנים, שפת גוף ותנועות עיניים. מחקרים מצבאים על כך שרגשות חיוביים, כמו סקרנות ושמחה, משפרים את יכולת הריכוז ואת יכולות הלמידה, בעוד שרגשות שליליים, כגון תשוכן וstress, עשויים לפגוע בתהליכי הקಲיטה של חומר הלימוד (Gupta et al., 2022).

מערכות לזיהוי רגשות מאפשרות ניטור של רמת המעורבות של תלמידים במהלך שיעור מקוון באמצעות ניתוח בזמן אמת הבאות הפנים ותנועות העיניים. דפוסים התנהגותיים מסוימים עשויים לשמש כאינדיקטור לרמת הקשב של התלמיד. למשל, תלמידים אשר מפניים את מבטם מהמסך בתדירות גבוהה או מפיגינים הבאות פנים חרשות ונשווים להיות במצב של חוסר ריכוז, שעומם או תשוכן. מודלים מבוססי רשתות נוירוניים קונבולוציוניות (CNNs) מסייעים בזיהוי דפוסים אלו, תוך חישוב מدد מעורבות המאפשר למורה לקבל מושב זמן אמיתי. באמצעות עיבוד מתקדם של התמונה והפקת מאפיינים חזותיים, המערכת יכולה להתריע על ירידת במעורבות ולהתאים את תכני השיעור לרמת הריכוז של התלמידים (Gupta et al., 2022).



איור 1 : זיהוי תחושים התלמידים בשיעור (Gupta et al., 2022).

יישום טכנולוגיות לזיהוי רגשות בלמידה מקוונת מתבצע באמצעות מספר גישות מרכזיות, אשר נבדלות זו מזו בדיקת הזיהוי, בדרישות החישוביות וביכולת העבודה בזמן אמיתי. הגישה הראשונה מבוססת על ניתוח הבעות פנים בלבד (Facial Emotion Recognition - FER), ובها נעשה שימוש במערכות למידה עמוקה, כגון CNN, לזיהוי רגשות ישירות מהבעות הפנים של התלמידים. יתרונה המרכזי של שיטה זו הוא פשוטות

היישום ויכולת העבודה בזמן אמת, אך היא רגישה לשינויים בתנאי התאורה ובזווית הצילום ועלולה שלא להבחן בין חוסר עניין לבין ריכוז גבוה.

גישה נוספת משלבת בין ניתוח הבעות פנים לבין מעקב אחר תנועות עיניים וشفת גוף, במטרה לספק ניתוח מكيف יותר של רמת המעורבות של התלמידים. באמצעות שילוב חיישני מעקב עיניים עם רשותות נוירונים, ניתן לזהות מצבים שבהם תלמיד נראה מרוכז לכואה אך בפועל דעתו מושחת. גישה זו מדויקת יותר, אך דורשת חומרה מתקדמת, כגון מצלמות בעלות דיווק גבוהה וחישוני מעקב תנועה, מה שמייקר את היישום ומקשה על שימוש רחב היקף (Pereira et al., 2024).

גישה שלישית, מתקדמת יותר, עשויה שימוש במודלים מולטי-מודליים (Multi-Modal Learning), המשלבים בין ניתוח תמונה, עיבוד דיבור (Speech Emotion Recognition - SER) ועיבוד שפה טבעית (NLP). שילוב זה מאפשר הבנה עמוקה יותר של הרגשות המובעים על ידי התלמידים, שכן הוא מתחשב לא רק בהבעות הפנים שלהם, אלא גם באופן הדיבור שלהם ובתוכן הנאמר. גישה זו מדויקת במיוחד, אך דורשת כוח חישובי משמעותי וזמן עיבוד ארוך יותר, וכן אינה מתאימה בהכרח ליישומים בזמן אמת.

אחד היתרוןות המרכזיים של מערכות לזיהוי רגשות הוא יכולת להתאים את החומר הלימודי לכל תלמיד. תלמידים המביעים הבעות של בלבול או תשכול יכולים לקבל הסברים נוספים, בעוד תלמידים שמבינים שמחה והבנה יכולים להתקדם מהר יותר בתהליכי הלמידה. מערכות אלו מסייעות גם בהוראה לתלמידים עם צרכים מיוחדים, שכן הן מאפשרות זיהוי מוקדם של תשכול או מצוקה, ובכך מספקות למורה כלים להתקנת השיעור לצרכים הייחודיים של התלמידים (Gupta et al., 2022).

למרות היתרוןות הרבים שמערכות אלו מציעות, קיימים אתגרים משמעותיים שיש לנקוט בחשבו. דיווק המודלים תלוי במגוון הנתונים שעליהם הם מאומנים, ולכן קיימת חשיבות לאימון על מאגרי נתונים רחבים הכוללים שונות רחבה של גזעים, תרבויות ותנאי תאורה. בנוסף, שמירה על פרטיות התלמידים היא סוגיה מרכזית, שכן מערכות אלו משתמשות במלומות למעקב בזמן אמת. יש צורך במידיניות פרטיות קפדנית וביישום של טכניקות כמו אונונימיזציה של נתונים או עיבוד מקומי על גבי המחשב האישי של התלמיד, במקרה של ליחת הנתונים לען (Pereira et al., 2024).

טכנולוגיות לזיהוי רגשות בזמן אמת מחייבת פדגוגית בימידה מקוונת בכך שהן מאפשרות מדידת מעורבות, התאמת חוות הלמידה ושיפור שיטות ההוראה באמצעות מידע אובייקטיבי. ההשוואה בין השיטות השונות מראה כי הבחירה בגישה תלולה במטרת היישום. עבור זיהוי רגשות בסיסי, ניתן להשתמש בעיבוד תמונה בלבד, ואילו עבור יישומים הדורשים רמת דיווק גבוהה יותר, ניתן לשלב מעקב עיניים או ניתוח דיבור. יישום מוצלח של טכנולוגיה זו מחייב התאמת אישית לצורכי התלמידים, Gupta et al., 2022; Jagadeesh & Baranidharan, 2022.

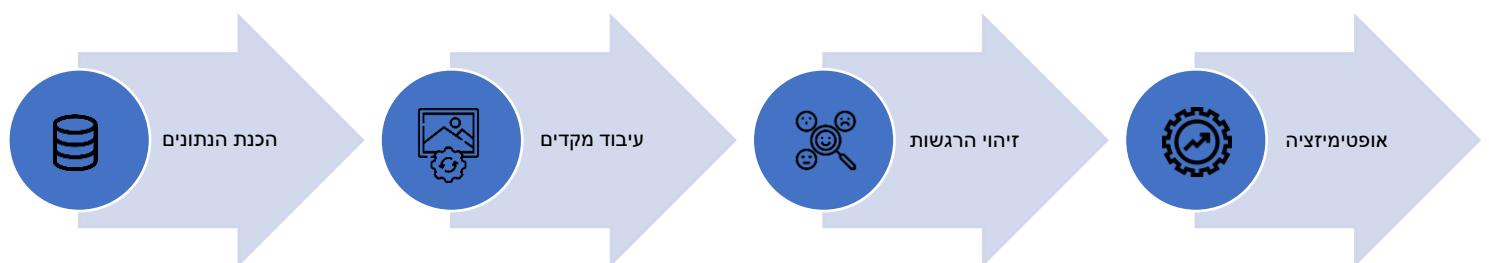
### **פרק 3 - מודלים וטכניקות בלמידה عمוקה לזיהוי רגשות**

פרק זה מותמקד בסקירה מעמיקה של מודלים متقدמים בלמידה عمוקה המשמשים לזיהוי רגשות מתוך הבעות פנים. תהליך הזיהוי נשען על טכניקות מתقدמות המשלבות עיבוד תמונה, חילוץ מאפיינים, ואימון רשתות נוירוניים כונבולוציוניות (CNN) לצד מודלים נוספים. הפרק מציג את התהליך החלם של בניית מערכת לזיהוי רגשות, מהכנת הנתונים ועד לשיפור ביצועי המודלים.

במהלך הפרק נסקור תהליכיים מרכזיים הכוללים :

1. הכנת הנתונים ועיבוד מקדים - סקירת מאגרי נתונים בולטים, הגדלת מאגרי נתונים באמצעות טכניקות Data Augmentation, עיבוד תמונה, זיהוי פנים, תיקון תמונות וניקוי רעש.
2. בניית הרשתות הנוירונית לזיהוי רגשות - הסבר על רשתות נוירוניים כונבולוציוניות (CNN), כולל הארכיטקטורה שלhn, תפקידי השכבות השונות ומימוש רשת כזו לצורך זיהוי רגשות.
3. סקירת מודלים בלמידה عمוקה לזיהוי רגשות - הצגת מודלים מסורתיים לצד מודלים מתقدמים, והשוואה ביניהם במונחי ביצועים.
4. אופטימיזציה ושיפור ביצועים - סקירת טכניקת Walrus לאופטימיזציה שמטרת להאיץ את זמן האימון ולשפר את הדיקוק של המודלים.

מטרת הפרק היא לספק תמונה רחבה ועמוקה על הגישות והשיטות הקיימות לזיהוי רגשות באמצעות למידה عمוקה, תוך התייחסות לאתגרים טכניים מרכזיים ולישומים המעשיים של התחום.



איור 2 : תהליך אפשרי לזיהוי רגשות.

#### **a. הכנות הנתונים ועיבוד מקדים**

##### **מאגרי נתונים**

מאגרי הנתונים בזיהוי רגשות מהווים את הבסיס לפיתוח ולמידה של מודלים בלמידה عمוקה בתחום זיהוי רגשות. הם מספקים דוגמאות לתמונות המכילות הבעות פנים שונות, אשר משמשות לאימון, בדיקה והערכתה של ביצועי המודלים. איקות מאגרי הנתונים, גודלם, ומגוון הדוגמאות שבהם משפיעים באופן ישיר על הדיקוק והכלליות של המודלים.

במיוחד בתחום זיהוי רגשות, מאגרי הנתונים חייבים לכלול מגוון רחב של הבעות פנים, זוויות צילום, תנאים תאורה, מגדרים, וגזעים, על מנת להבטיח שהמודלים יהיו עמידים במצבים אמיתיים. יתרה מזאת, תיוג מדויק של הרגשות במאגרים הוא קריטי, שכן אי-דיוקים בתוצאות עלולים להוביל לתוצאות שאגויות במהלך האימון.

במהלך השנים פותחו מספר מאגרי נתונים פופולריים המשמשים חוקרם בתחום זה. המאגר הראשון (FER-2013 - Facial Expression Recognition) הוצג על תחרות Kaggle ב-2013. המאגר כולל כ-35,000 תמונות בגווני אפור של פיצופים בגודל 48x48 פיקסלים. התמונות מתוויות לשבע קטגוריות של רגשות: שמחה, עצב, כאס, פחד, הפתעה, גועל וניטרלי (Pereira et al., 2024). מאגר זה נפוץ בזכות זמיןותו הפומבית וגודלו, אך הוא מוגבל בכך שההבעות שבו מתועדות בתנאי צילום מבודדים יחסית. מאוחר יותר פותח CK+ (Extended Cohn-Kanade Dataset) המאגר נחשב מדויק להבאה מוגדרת. המאגר מכסה שבע קטגוריות של רגשות בדומה ל-FER-2013. המאגר נחשב מדויק מאוד אך מוגבל בגודלו ובמגוון הקטגוריות (Joseph et al., 2024).

RAF-DB (Real-world Affective Faces Database) כולל מעל 30,000 תמונות פנים, מתוויות לפי רגשות עיקריים ומשניים. המאגר מתעד הבעות פנים בתנאי תאורה טבעיים ולא מבודדים, מה שהופך אותו לשימושי עבור יישומים במצבים ריאליים. המאגר משלב תוצאות שנעשו על ידי מספר מתאגידים כדי להבטיח דיוק גבוה (Gupta, Kumar & Tekchandani, 2023).

WIDER FACE כולל מעל 32,000 תמונות, עם יותר מ-393,000 פיצופים מתויגים. המאגר מכסה מגוון רחב של פרמטרים, כולל זוויות צילום, תנאים תאורה, והבעות פנים. משמש בעיקר לזיהוי פנים אך רלוונטי גם לזיהוי רגשות בשל המגוון הרב (Gupta, Kumar & Tekchandani, 2023).

Database	Posed/ In the Wild	Images	Type	Expressions
FER-2013	In the wild	35,000	gray	anger, disgust, fear, happiness, sadness, surprise, neutral
CK+	posed	593	Mostly gray	neutral, sadness, surprise, happiness, fear, anger, contempt, disgust
RAF-DB	In the wild	29,000	color	anger, disgust, fear, happiness, sadness, surprise, neutral

איור 3 : טבלת השוואה בין מאגרי המידע המוצגים (Pereira et al., 2024)

#### אתגרים בעבודה עם מאגרי נתונים:

- מגוון מוגבל - מאגרי נתונים מסוימים אינם כוללים מגוון רחב מספק של משתנים כמו גיל, גזע ותנאי צילום, מה שモוביל למודלים מוטיבים.
- תיוג לא מדויק - במאגרים גדולים, טוויות תיוג הן נפוצות ועשויות להשפיע על הביצועים של המודלים.
- גודל מוגבל - מאגרי נתונים קטנים מדי עשויים להוביל לאימון לא מספק של מודלים, במיוחד כאשר משתמשים בראשות עמוקות עם מספר רב של פרמטרים.
- צילומים מבודדים מדי - רבים ממאגרי הנתונים צולמו בתנאים מבודדים מאוד, ולכן המודלים המיועדים לזיהוי רגשות בתנאי שטח מתקשים להתמודד עם רעשים וזריזות משתנות.

- שונות תרבותית - הבעיות פנימית עשוות להתפרש בצורה שונה בתרבותות שונות, אך מרבית המאגרים מתמקדים בקבוצות אוכלוסייה ספציפיות.

מודלים בלמידה عمוקה דורשים כמות משמעותית של נתונים על מנת ללמידה דפוסים מורכבים ולהציג ביצועים מדויקים. עם זאת, מאגרי נתונים רבים בתחום זיהוי רגשות מוגבלים בגודלם ובמגוון שלהם. בעיה זו יכולה להוביל לאימון לא מספק של המודלים ולהטויות בתוצאות, במיוחד כאשר הנתונים אינם מייצגים את המציאות באופן רחב ו מגוון. כדי להתמודד עם אתגר זה, נעשה שימוש בטכניקה הנקראת **Data Augmentation** - הגדלת מאגרי נתונים מלאכותית.

**Data Augmentation** מאפשרת להתגבר על מוגבלות גודל המאגר על ידי יצירת דוגמאות חדשות מהנתונים המקוריים, ובכך להגדיל את הגיון במאגר ולהפחית את הסיכון לאימון יתר (Overfitting). תהליך זה חשוב במיוחד במאגרים שבהם יש ייצוג מוגבל של קטגוריות רגשות, תנאי תארה משתנים, או זוויות צילום לא סטנדרטיות. בנוסף, הגדלת המאגר מסייעת להפוך את המודלים ליעדים יותר לרעים ולתנאים מציאותיים, כפי שמודגם במחקריהם נספים על הנתונים המקוריים, במטרה להרחיב את מאגרי הנתונים ולשפר את הביצועים של המודלים. אחת הטכניקות המרכזיות היא שימוש בטרנספורמציות גיאומטריות, שבחנו ניתנת לביצוע סיבוב (Rotation) של התמונות בזווית קטנות, כגון  $15 \pm$  מעלות, כדי לדמות זוויות צילום שונות. בנוסף, ניתן לישם חיתוך (Cropping) של התמונות לחלקים קטנים יותר תוך שמירה על המידע החיוויי (Pereira et al., 2024), או שינוי קנה מידת (Scaling) שמאפשר הגדלה או הקטנה של האובייקט בתמונה (Jagadeesh, 2022).

שינויים בתארה מהווים עוד גישה חשובה ב- **Data Augmentation**. לדוגמה, התאמת בהירות (Brightness Adjustment) משמשת לשינוי עצמת התארה בתמונה, בעודו הוספת רעש (Noise) משוחרת תנאי צילום גרעינים על ידי הוספת רעש אקריאי. בנוסף, קיימות טכניקות לשינוי הצבע בתמונה. שינוי רמת הצבעים (Color Jittering) מבוצע על ידי שינוי רנדומלי של גווני הצבע בתמונה, והמרה לגווני אפור (Grayscale Conversion) הופכת תמונה צבעונית לשחור-לבן, מה שעשו לטיען למודלים להתמקד במאפיינים מבניים ולא בצבעים.

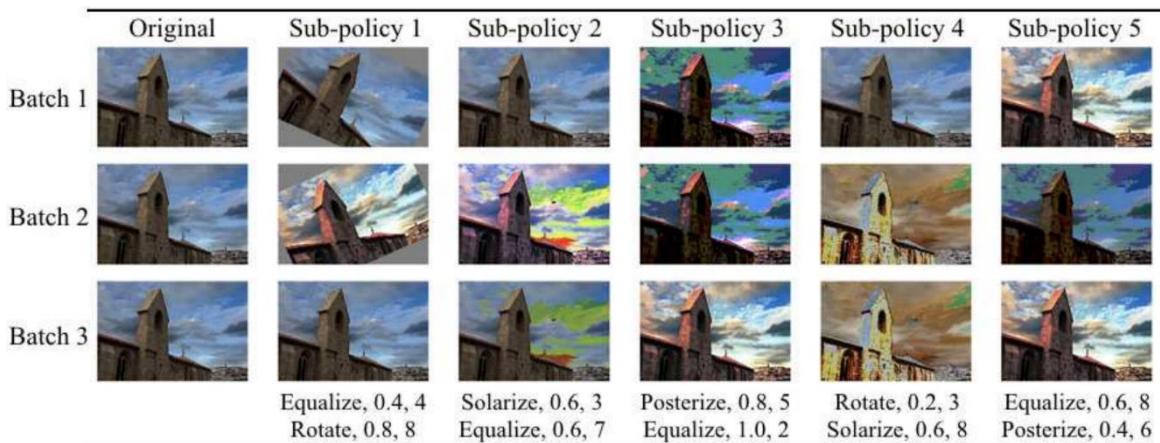
אלגוריתמים רבים פותחו כדי להקל על תהליך ה- **Data Augmentation** ולבצע אותו באופן אוטומטי :

- **Augmentor** - ספרייה פופולרית המאפשרת לבצע מגוון טרנספורמציות על תמונות (Pereira et al., 2024)
- **AutoAugment** - אלגוריתם מבוסס למידה عمוקה אשר מזהה אוטומטית את השימוש האופטימלי של טרנספורמציות עבור מאגר נתונים מסוים.
- **GANs (Generative Adversarial Networks)** - רשות מתקדמות المسؤولות ליצור תמונות חדשות לחלוין המבוססות על מאגר הנתונים המקוריים (Pereira et al., 2024).

אחד היתרונות המרכזיים של **Data Augmentation** הוא שיפור הביצועים של המודלים, שכן הגדלת המאגר מסייעת לשפר את הדיק ו את הכלליות של המודלים בזיהוי דפוסים חדשים. בנוסף, הטכניקה

מאפשרת ייצור עמידות גבוהה יותר לרעש ותנאי צילום משתנים, דבר ההופך את המודלים למתאימים יותר לשימוש במצבים מציאותיים. יתרון נוסף הוא החיסכון בעלות, לאחר ואין צורך באיסוף נתונים חדשים, אלא ניתן להשתמש בתנאים קיימים ולבצע עליהם מניפולציות מגוונות.

עם זאת, ל - Data Augmentation יש גם חסרונות. מגבלה עיקרית היא היצירתיות המוגבלת, שכן הטכניקות מתבססות על הנתונים המקוריים ואין אפשרות ליצור חדשים או נתונים שאינם קיימים המקורי. בנוסף, שימוש ביוטר מדיה טרנספורמציות עשוי להוביל למורכבות בתהיליך האימון ולהאטת שמעותית ביצועיו. יתרה מזאת, אם התהיליך אינו מבוקר כראוי, קיימת סכנה להdagשת הטוויות קיימות במאגר הנתונים, מה שעלול להשפיע לרעה על דיקוק המודלים ולשמור הטוויות לא מכוונות.



.(Cubuk et al., 2022) Data Augmentation

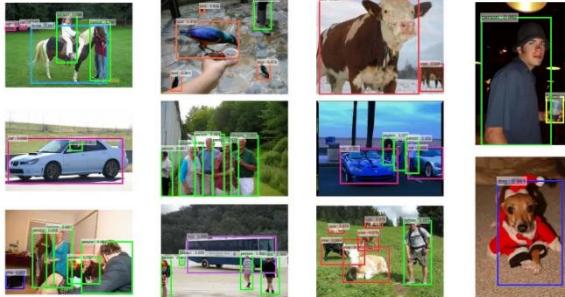
#### זיהוי הפנים - כלים וטכניקות

זיהוי פנים מהו זה שלב קריטי בתהיליך זיהוי רשות באמצעות מידת עומקה. זיהוי מדויק של אזור הפנים בתמונה מאפשר למודל להתמקד באזוריים הרלוונטיים, לחץ מאפיינים משמעותיים, ולשפר את הדיקוק הכלול של המערכת. תהליך זה כולל שימוש בכלים וטכניקות מתקדמות מותאמות לתנאי צילום משתנים, כגון זווית צילום שונה, תאורה בלתי אחידה, והבעות פנים חלקיות.

אחד השיטות הנפוצות לזיהוי פנים היא השימוש באלגוריתם Viola-Jones, שמציע ביצועים מהירים במיוחד בתנאים מבוקרים. אלגוריתם זה פועל על ידי חלוקת התמונה לאזוריים קטנים ונינוחים לזיהוי אזוריים הדומים למאפייני פנים. עם זאת, השיטה פחותה יعلا במצבים שבהם הפנים מצלומת בזוויות לא שגרתיות או בתנאי תאורה ירודים (Joseph et al., 2024).

אלגוריתמים מודרניים יותר, כמו, Faster R-CNN מציעים פתרון מתקדם לזיהוי פנים בתנאים ריאליים. Faster R-CNN משלב זיהוי אזוריים רלוונטיים יחד עם סיוגם, ובכך מאפשר זיהוי פנים מהיר ומדויק גם בזוויות צילום לא שגרתיות או בתנאי תאורה טבעיות. כלי זה נחשב לאחד הייעילים ביותר ביום לזיהוי פנים לזיהוי רשות (Pereira et al., 2024).

אחד האתגרים המשמעותיים בזיהוי פנים הוא התמודדות עם רעש רקע בתמונה, במיוחד כאשר הפנים אינם האלמנט המרכזי בתמונה. בעיה זו עלולה לגרום לזיוהוי שגוי או לפספוס פנים בתמונה. טכניקות כמו הוספת טרנספורמציות לרקע בזמן האימון משפרות את עמידות המודל[Pereira et al., 2016]. (2024).



איור 5 : תוצאות שימוש ב

.(Ren et al., 2016)

#### תיקון תמונות ונקיי רעש

ביעות כמו רעש דיגיטלי, תנאי תאורה לא אחידים ועיוותים בתמונה עלולות לגרום לזיוהוי שגוי של רגשות, ולכן יש צורך בשיטות תיקון מתקדמות על מנת להבטיח שהנתונים יהיו נקיים, עקובים ומתאים לאימון.

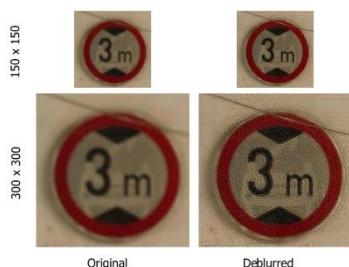
רעש בתמונה נוצר עקב גורמים שונים, כמו רזולוציה נמוכה של מצלמה, דחיסת תמונה אגרסיבית או הפרעות בתנאי הצילום, והוא עלול להשפיע על זיהוי הבעיות פנים. מסנן Gaussian Blur הוא אחת השיטות הנפוצות להסרת רעשים על ידי חילוק התמונה תוך שמירה על מאפיינים מרכזיים. בינו לבין המסננים ליניאריים פשוטים, מסנן גאוסיאני משתמש בהתפלגות נורמלית כדי לתת משקל שונה לכל פיקסל בסביבתו, וכך הוא מפחית את הרעש מבלי לטשטש קווים חזקים או פרטים חשובים בתמונה. המסנן פועל על ידי קונволוציה של התמונה עם גרעין (Kernel) המוגדר על ידי ההתפלגות הגaussיאנית :

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

כasher :

- $G(x, y)$  מייצג את פונקציית ההתפלגות של הפילטר.
- $\sigma$  הוא סטיית התקן של ההתפלגות, אשר קובע את רמת החילקה.
- $x, y$  הם הקואורדינטות של הפיקסלים.

כל שהערך של  $\sigma$  גדול, כך התמונה תוכלך יותר, אך ייתכן ואובדן מידע יתרחש. לכן, נבחר ערך מסוון של  $\sigma$  כך שהרעש יופחת אך המבנה של הפנים ישמר [Liu et al., 2020].



איור 6 : תוצאות שימוש ב

.(Liu et al., 2020)

שינויי תוארה מהווים גורם משמעותי בהשפעה על זיהוי רגשות בתמונות פנים. תמונות שצולמו בתנאי תאורה נזוקים או בעלי ניגודיות חלשה עלולות לגרום למודלים לטעות בזיהוי תווים פנים. שיטת Histogram Equalization (השוואת היסטוגרמה) נועדה לשפר את הניגודיות בתמונה על ידי פיזור טוב יותר של רמות הבחרות.

ההיסטוגרמה של תמונה מייצגת את התפלגות ערכי הבהירות (intensity values) של הפיקסלים, כאשר בתמונות כהות רבים מהפיקסלים מרכזים בטוחים נזוקים, ואילו בתמונות בהירות הפיקסלים נוטים לטוחים גבוהים. המטרה של Histogram Equalization היא למתוח את ערכי הבהירות כך שייפרסו על פני כל הטווח האפשרי, מה שմষפר את הבדיקה בין פרטיהם לבין עצמם.

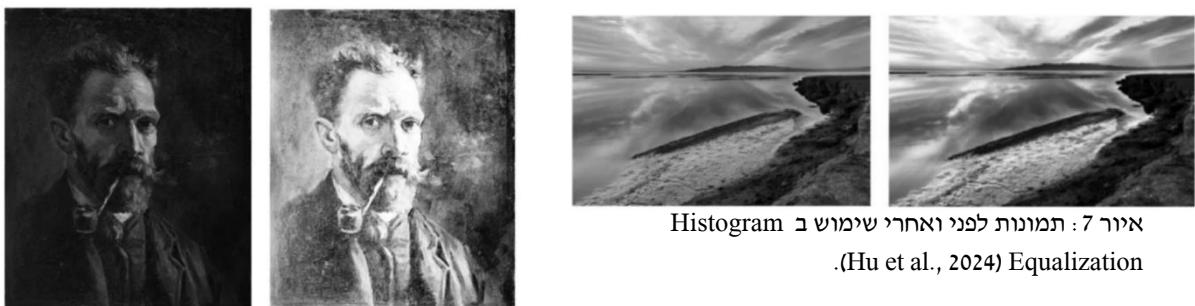
תהליך תיקון התוארה מבוצע על ידי חילוף פונקציה מצטברת על ההיסטוגרמה של התמונה:

$$T(i) = \frac{\sum_{j=0}^i h(j)}{N}$$

כאשר :

- $i$  הוא הערך המתווך של הבהירות עבור דרגת עוצמה  $i$ .
- $j$  הוא מספר הפיקסלים עם ערך בהירות  $j$ .
- $N$  הוא מספר הפיקסלים הכללי בתמונה.

שיטת זו מתאימה במיוחד לשיפור פרטים בפנים. שכן היא מאפשרת למודל להבדיל טוב יותר בין אזורים מוצלים לאזורים מוארים. עם זאת, היא אינה יעילה כאשר התוארה בתמונה אינה אחידה, ולכן לעיתים נעשה שימוש בטכניקה מתקדמת יותר, (AHE), Adaptive Histogram Equalization, המפרקת את התמונה לאזורים קטנים יותר ומבצעת השוואת היסטוגרמה לכל אזור בנפרד (Hu et al., 2024)



איור 7 : תמונות לפני ואחרי שימוש ב Histogram Equalization .(Hu et al., 2024)

שיפור הניגודיות באמצעות Histogram Equalization מסייע בהגברת בהירות התמונה ושיפור הניגודיות, ובכך מאפשר הפקת מידע חזותי עשיר יותר, במיוחד בתמונות בעלות תאורה ירודה. מחקרים הראו כי Histogram Equalization מגביר את האנטרופיה של התמונה, משפר את התפלגות הבחרות ומסייע בשימור פרטים חיוניים באזוריים מוצללים.

נמצא כי שיפורים מתקדמים של האלגוריתם, כמו RCLAHE, משפרים אף יותר את אינכות התמונה על ידי התאמת מקומית של הניגודיות. ניסויים הראו כי שיטה זו הגדילה את ממד האנטרופיה ב-12%, מה שomid על שימור טקסטורה טובה יותר ושיפור איחודות התוארה (Hu et al., 2024).

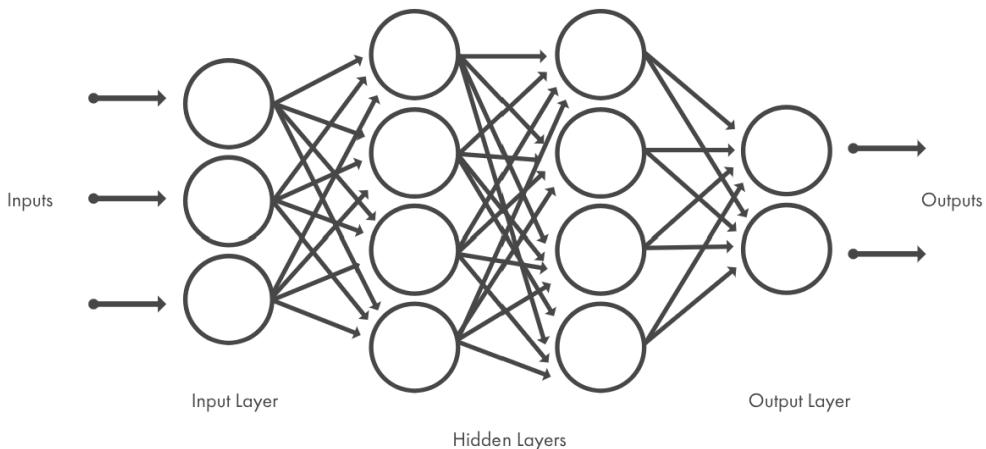
## **ב. זיהוי רגשות**

רשתות נוירוניים קוונטולוציוניות (CNN - Convolutional Neural Networks) הן אחת הטכנולוגיות המרכזיות בתחום למידה عمוקה וריאית מחשב. תחילת התפתחותן הייתה בשנות ה-80, אך רק עם העלייה בכוח החישוב והופעת מאגרי נתונים גדולים הן הצליחו להגיעה לביצועים ברמה על-אנושית במשימות עיבוד תמונה.

פריצת דרך משמעותית חלה עם פיתוח LeNet-5 (Michalski & Plechawska-Wójcik, 2022) שסייע לשינויים זיהוי ספרות כתובות, ובהמשך, רשתות כמו VGG, AlexNet, ResNet-שיפרו משמעותית את הביצועים בתחום.

רשתות נוירוניים הן מודלים מתמטיים המבוססים על מבנה המוח האנושי, المسؤولים ללמידה דפוסים מותוך נתונים. ביגוד לרשתות נוירוניים רגילות (Fully Connected Networks), שבו מסוף הפרמטרים גדול באופן משמעותי ככל שהקלט גדול, CNN נבנו במיוחד כדי לנתח מידע מרחבית כמו תמונות, תוך שימירה על יחסים מרחביים בין הפיקסלים.

CNN פועלות באמצעות פילטרים (Kernels) החולפים על פני התמונה ומשהים דפוסים כמו קווים מתאר, טקסטורות וצורות גיאומטריות. תהליך זה נקרא קוונטולציה, והוא מאפשר לרשת לזהות חלקים חשובים בתמונה. לאחר שלב הקוונטולציה, מתקבצת הफחתת ממדים (Pooling) כדי לשמר את המידע הקרייטי תוך הפחיתת כמות החישובים. לבסוף, שכבות Fully Connected מבצעות סיווג סופי של התמונה.



אирור 8 : ארכיטקטורה כללית של רשת נוירוניים [8].

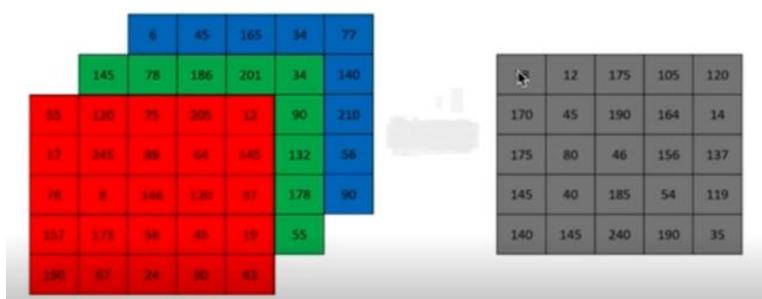
רשתות CNN מתאימות במיוחד למשימות של זיהוי רגשות, שכן הן מסוגלות לחץ דפוסים עדים נחים בהבעות פנים ולזהות תזוזות קטנות בעיניהם, בגבורה ובפה. תכונות אלו מאפשרות למודלים לנתח שינויים רגשיים עדים ולהתאים עצם לתנאי צילום שונים באמצעות Data Augmentation, שמאפשר ייצור גרסאות שונות של התמונות להגדלת הגיון במאגר הנתונים.

## ארQUITטורה של CNN

### 1. שכבת הקלט (Input layer)

שכבה הקלט היא השלב הראשון בתהליך עיבוד התמונה בראשת נוירונים קונבולוציונית. היא מקבלת תמונה ומיירה אותה לייצוג מספרי המאפשר לרשף לנתח את המידע ולהפיק ממנו תובנות. בשכבה זו לא מתבצע עיבוד מתקדם, אך היא קריטית בהבטחת פורמט נתונים אחיד והקמת הקלט כך שיתאים למודל. רשותות CNN עשוות שימוש בתמונות בגווני אפור או בצבע (RGB), ולכן שכבת הקלט חייבה להתאים עצמה לסוג הקלט שהמודל מתוכנן לטפל בו (Krizhevsky et al., 2012).

כאשר הרשות מקבלת תמונה בגווני אפור, היא מייצגת כל פיקסל כערך יחיד בין 0 ל-255, כאשר 0 מסמל צבע שחור ו-255 מייצג צבע לבן. מבנה זה מיוצג כמטריצה דו-ממדית בגודל  $W \times H$  שבה  $H$  מסמן את גובה התמונה ו- $W$  את רוחבה. לעומת זאת, בתמונות צבעוניות, כל פיקסל מכיל שלושה ערויצים - אדום, ירוק וכחול (RGB), כך שהתמונה כוללת מייצגת כמטריצה תלת-ממדית בגודל  $3 \times W \times H$ . לדוגמה, תמונה בגודל  $48 \times 48$  פיקסלים בגווני אפור מתוארת כמטריצה  $X \in R^{48 \times 48}$  ואילו תמונה צבעונית באותו גודל מייצגת כ- $X \in R^{48 \times 48 \times 3}$ .



איור 9 : המחשה של תמונה מסווג  $(5 \times 5 \times 3)$  RGB ומשוג  $(5 \times 5 \times 1)$  Grayscale.[15]

לפני שהתמונה מועברת לעיבוד בראשת הנוירונים, היא עוברת מספר שלבים של עיבוד מקדים, שנועד לשפר את איזoct הנ נתונים ולהבטיח שהמידע שיוזן למודל יהיה אחיד ונטול הפרעות. שלבים אלה כוללים שינוי גודל, נרמול והסרת רעש, אשר מסייעים להפחית מורכבות החישובים ולמניעת הטיאות שעלולות לפגוע בבדיקה התחזיות של המודל.

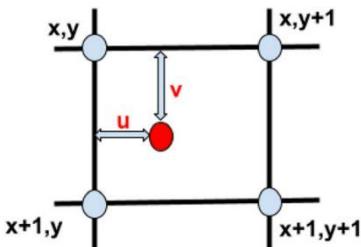
אחד מהשלבים הראשונים בעיבוד המקדים הוא שינוי גודל התמונה (Resizing), כדי להתאים אותה לדרישות המודל. לאחר שמודלים כמו VGG-16 דורשים תמונות בגודל  $224 \times 224$  פיקסלים, בעוד שמאגר כמו FER-2013 לדוגמה מכיל תמונות בגודל  $48 \times 48$  פיקסלים, יש לבצע אינטראפולציה כדי לשנות את הגודל תוך שמירה על פרטי התמונה (Rathee et al., 2021).

באינטרפולציה ביליניארית לדוגמה, מחושב הממוצע המשוקל של ארבעת הפיקסלים הקרובים ביותר, והنتוצאה נחשבת לערך החדש של הפיקסל. קביעת ערך הבחרות עבור הפיקסל החדש מתבצעת לפי הנוסחה הבאה:

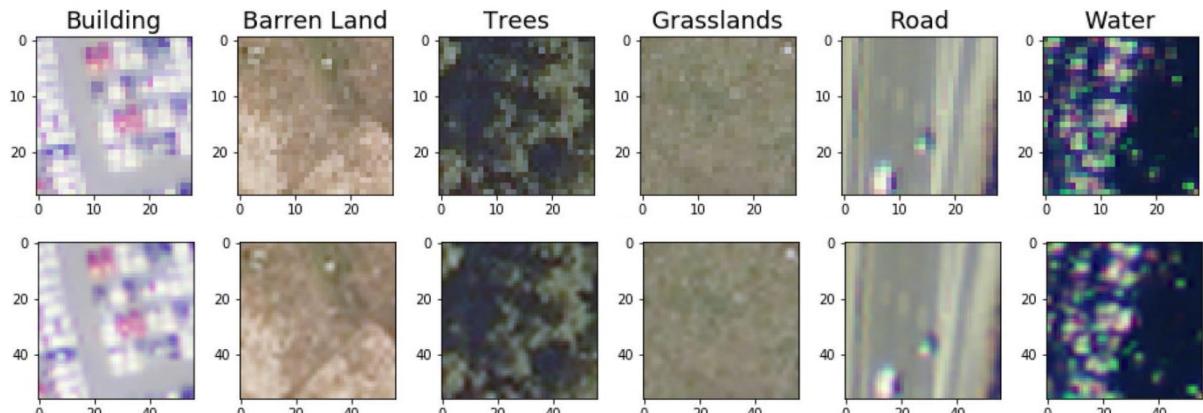
$$P_{target}(x_0, y_0)$$

$$= (1-u)(1-v)P_{source}(x, y) + u(1-v)P_{source}(x, y+1) \\ + (1-v)vP_{source}(x+1, y) + uvP_{source}(x+1, y+1)$$

כאשר  $(x_0, y_0)$  הוא ערך הפיקסל בתמונה המעובדת,  $P_{source}(x, y)$  הוא ערך הפיקסל בתמונה המקורית ו- $u, v$  מייצגים את המרחקים האופקיים והאנכיאים של הנקודה שmoved מהפינה הראשונה  $(x, y)$ , כפי שמודגס באירוע 10 (Rathee et al., 2021).



אייר 10 : ייצוג סכמטי עבור השכן הבילינארי הקרוב ביותר (Rathee et al., 2021).



אייר 11 : המחשה של שימוש באינטראפלציהbilinear בילינארית (Rathee et al., 2021).

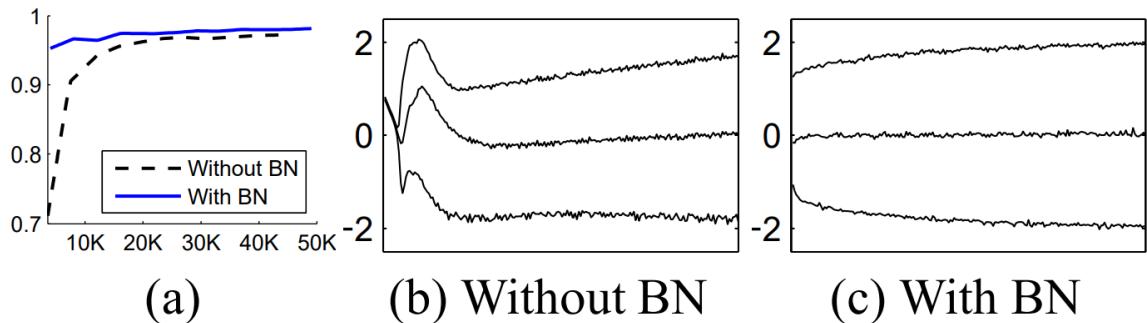
שלב חשוב נוסף הוא נרמול הנתונים (Normalization), שמטרתו להביא את ערכי הפיקסלים לטווח נוח לעיבוד נוירוניים ולמנוע חוסר יציבות במהלך האימון. נרמול סטנדרטי נעשה באמצעות חישור הממוצע  $\mu$  וחלוקת בסטיית התקן  $\sigma$ :

$$\frac{X - \mu}{\sigma} = X_{norm}$$

גישה נוספת היא חלוקת הערכים ב-255 כדי להביא את הפיקסלים לטווח שבין 0 ל-1:

$$\frac{X}{255} = X_{norm}$$

נرمול זה מסייע להתקנסות מהירה של הרשת ומונע שגיאות חישוביות שעוללות להתרחש כאשר הנתונים גדולים מדי (Ioffe, 2015).



איור 12 : ממחיש את השפעת Batch Normalization על אימון רשת נוירונים.

חלק (a) : הרשת מתכנסת מהר יותר ומשינה דיקט גובה יותר עם .Batch Normalization חלקים (b, c) : ההתפלגות של הקלטים לרשת יציבה יותר במהלך האימון, מה שמחזית Internal Covariate Shift ומשפר את יציבות הלמידה (Ioffe, 2015).

במהלך האימון, התמונות לא מוגנות לרשות בזו אחר זו, אלא בקבוצות הנקראות אצודות (Batches), אשר משפרות את ביצוע החישוב ואת יציבות הלמידה. כל אצודה מכילה מספר תמונות, ובנה הנתונים נשמר במערך ארבע-ממדי שבו B מייצג את מספר התמונות בכל אצודה :

$$X_{batch} \in R^{B \times H \times W \times C}$$

לדוגמה, כאשר מבוצע אימון עם 32 תמונות בגודל  $48 \times 48$  בגוני אפור, הנתונים נשמרים כך :

$$X_{batch} \in R^{32 \times 48 \times 48 \times 1}$$

ואילו עבור מודל כמו ResNet-50, שעבוד עם תמונות צבעוניות בגודל  $224 \times 224$ , הקלט יהיה :

$$X_{batch} \in R^{32 \times 224 \times 224 \times 3}$$

שכבה הקלט מלאת תפקיד קרייטי בהכנות הנתונים לעיבוד ברשת CNN היא אחראית על המרת התמונה לפורתט מספרי, התאמת גודל הקלט כך שיתאים למודל, נרמול הנתונים למניעת חוסר יציבות במהלך האימון, וארגון התמונות באצודות כדי לשפר את יעילות החישוב. תהליכי אלו מבטיחים שהתמונה תנוטח באופן אופטימלי ומונעים שגיאות שעוללות להשפיע על תהליך הלמידה.

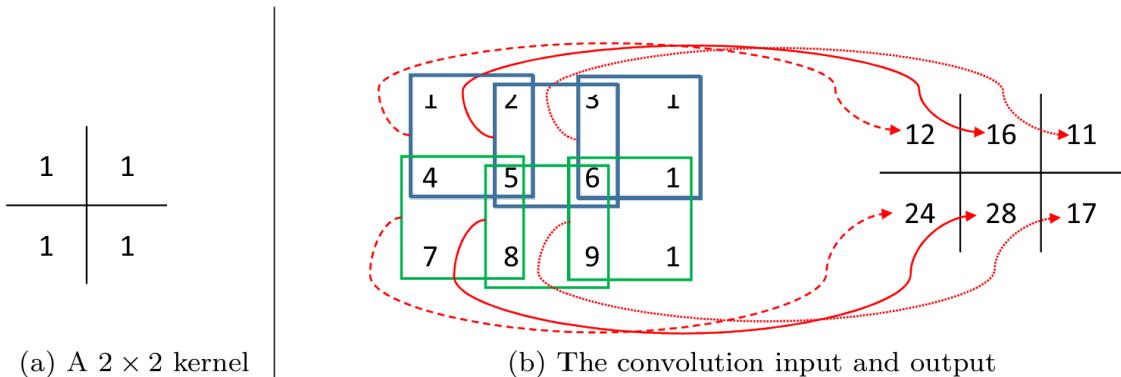
## 2. שכבות קונבולוציה (Convolutional Layers)

שכבה הקונבולוציה היא הליבה של רשתות נוירונים קונבולוציוניות, והיא אחראית על חילוץ מאפיינים (Feature Extraction) מתוך התמונה הקלט. השכבה פועלת על ידי החלת פילטרים (Kernels) על התמונה, המסייעים לזהות דפוסים חזותיים חזורים, כמו קצוטות, מרקמים, ועצמים מורכבים יותר. בשכבות הראשונות של הרשת, הפילטרים מזוהים מאפיינים בסיסיים כמו קוים אופקיים או אנכיים, ואילו בשכבות המאוחרות יותר, הם מזוהים מבנים מורכבים יותר, כגון חלקים של פנים ו הבעות רגשות (Wu, 2017).

התהליך המתמטי של הקונבולוציה מתבצע על ידי החלקת פילטר בגודל קבוע על פני התמונה, חישוב מכפלה פנימית בין ערכי הpixels לבין ערכי הפילטר, והפקת מפת תכונות (Feature Map) המיצגת את הדפוסים שנמצאו בתמונה. הפעולה מתבצעת לפי הנוסחה הבאה (Wu, 2017) :

$$y_{i^l+1, j^l+1, d} = \sum_{i=0}^H \sum_{j=0}^W \sum_{d^l=0}^{D^l} f_{i,j,d^l,d} \times x_{i^l+1+i, j^l+1+j, d^l}^l$$

כאשר  $f_{i,j,d^l,d}$  מייצג את ערכי הקלט של השכבה,  $f_{i,j,d^l,d}$  הוא הernal המופעל,  $y_{i^l+1, j^l+1, d}$  הוא הפלט של שכבת הקונבולוציה עבור הernal מסוים במקומות המתאים,  $H$  ו-  $W$  הם ממדיו הernal,  $D^l$  הוא מספר העורוצים בשכבת הernal ו-  $d$  מייצג את העורץ בפלט.



אייר 13 : המחשה של תהליך הקונבולוציה עם קernal  $2 \times 2$ . הריבועים הצבעוניים מסמנים את אזורי הקלט שעלייהם מוחל הernal, והחצים האדומים מראים כיצד כל אזור מומר לערכי הפלט במפת התכונות (Wu, 2017).

לשכבת הקונבולוציה יש מספר היפר-פרמטרים קריטיים המשפיעים על ביצועי המודל. אחד מהם הוא גודל הפילטר  $f$ , אשר קובע את גודל האזור שכל יחידה בשכבה הקונבולוציונית רואה בכל שלב. ערכים טיפוסיים הם  $3 \times 3$  או  $5 \times 5$ , כאשר פילטרים קטנים עשויים זיהוי פרטיים עצינים, ואילו פילטרים גדולים יותר מזהים מבנים רחבים יותר. פרמטר נוסף הוא הצעד (Stride) הקובע בכמה פיקסלים הפילטר מתקדם בכל שלב. ערך פסיעה של 1 שומר על רצולוציה גבוהה יותר, בעוד שערכים גבוהים יותר מביאים לצמצום גודל הפלט (Wu, 2017).

בנוסף, ניתן להוסיף Padding (ריפורוד) המשמש להוספת שורות ועמודות סביב התמונה לפני ביצוע הקונבולוציה, על מנת לשנות על גודל הפלט. ללא Padding, הגודל המרחב של הפלט קטן יותר מזה של הקלט כאשר גודל הernal גדול מ  $1 \times 1$ . עם זאת, לעיתים נדרש שהפלט יהיה בעל אותו ממדים כמו הקלט, וכן נעשה שימוש ב-Padding.

אם הפלט בגודל  $D^l \times W^l \times H^l$  והernal בגודל  $D \times W \times H$ , גודל הפלט ללא Padding מחושב כך :

$$H_{out} = H^l - H + 1, \quad W_{out} = W^l - W + 1$$

כדי לשמור על גודל הפלט זהה לקלט, יש להוסיף Padding כך שכל עורץ בклט מורחב על ידי הוספה

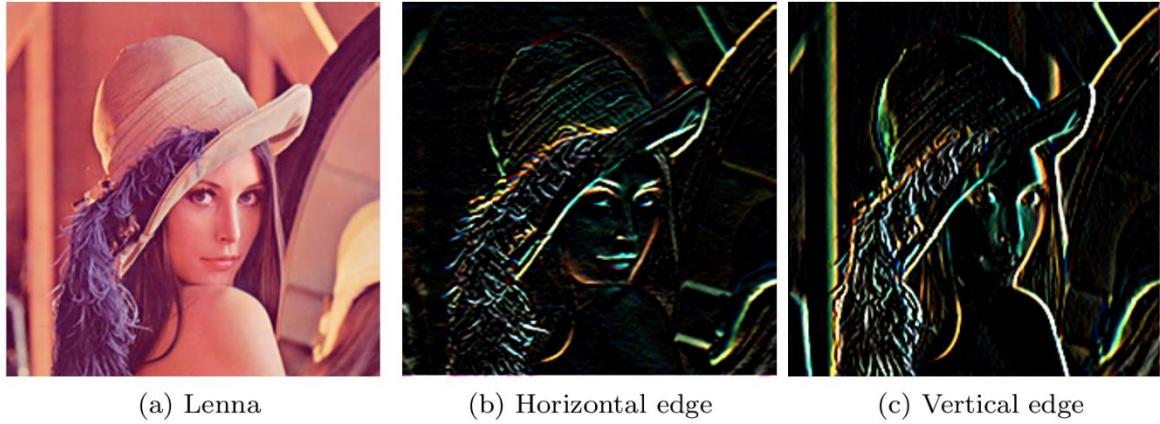
שורות מעלה השורה הראשונה ו-  $\left\lceil \frac{H-1}{2} \right\rceil$  שורות מתחת לשורה الأخيرة. באופן דומה, מוסיפים

עמודות משמאל לעמודה הראשונה ו-  $\left\lceil \frac{W-1}{2} \right\rceil$ עמודות מימין לעמודה الأخيرة.

פעולה זו מבטיחה שהממדים של הפלט נותרים  $D \times W \times H$ .

לרוב, הערכים המתווספים ב Padding הם אפסים, אך ייתכן גם שימוש בערכים אחרים.

לאחר ביצוע פעולת הקונволוציה, מתקבלת מפת פלט חדשה, שבה כל ערך מייצג את תוצאת הפעלת הקרNEL על אזור מסוים בקלט. כאשר משתמשים במספר קרNELים שונים באותה שכבה, ניתן להפיק מספר מפות פלט, כאשר כל קרNEL מזוהה דפוסים שונים בתמונה. לדוגמה, קרNEL מסוים עשוי לזהות קצוטות אופקיים, בעוד קרNEL אחר יתמקד בזיהוי קוויים אנכיים או טקסטורות מורכבות יותר.



איור 14 : מציג תמונה קלט צבעוני (a) ואת תוצאות הקונבולוציה שלה עם שני קרNELים שונים (b,c). הפעלת הקרNELים על תמונה מאפשרת ליזות קצוטות אופקיים (b) או קצוטות אנכיים (c), בהתאם לקרNEL שנבחר (Wu, 2017).

אחד היתרוןות המרכזיים של שכבת הקונבולוציה הוא שימור הקשר המרחבי בין הpixelsים תוך שמירה על ייעילות פרמטרית. במקרים שכל נוירון יהיה מחובר לכל פיקסל בקלט, כפי שקרה בראשת **Fully Connected**, הkerNELים פועלם באופן מקומי, מה שמאפשר למודול למדוד דפוסים אזוריים ולחולוק פרמטרים לאורך כל התמונה. עובדה זו מפחיתה משמעותית את מספר הפרמטרים, ומאפשרת לרשת להיות חסכונית יותר מבחןת משאבי חישוב.

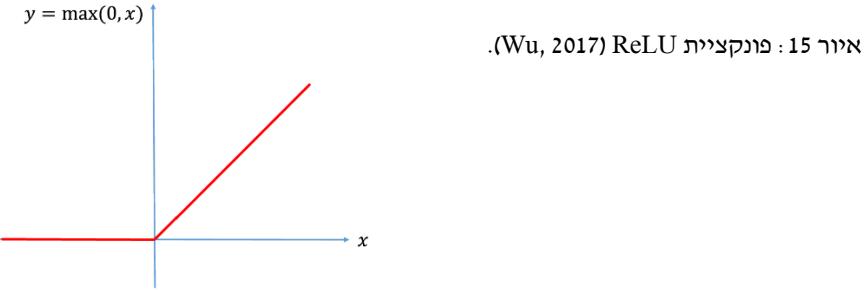
### 3. פונקציית האקטיבציה (Activation Function)

פונקציית האקטיבציה מאפשרת לרשת למדוד קשרים לא-lienאריים ולהבחין בין דפוסים מורכבים. לאחר שכל שכבת קונבולוציה מפיקה מפות תכונות(Feature Maps), יש צורך להכניס אי-liニアריות כדי לאפשר לרשת ליזות יחסים מורכבים בתמונה. ללא פונקציית האקטיבציה, הרשת הייתה מותנהגת כמודלlienארי פשוט, מה שהייתה מגבלת את יכולת ההסקה שלה (Wu, 2017).

הfonקציה הנפוצה ביותר היא ReLU (Rectified Linear Unit), המוגדרת כך :

$$f(x) = \max(0, x)$$

fonקציה זו מחליפה ערכים שליליים ב-0 ומשאירה ערכים חיוביים ללא שינוי. יתרון של ReLU הוא פשוטות חישובית ומניעת בעיות כמו דעיכת גרדיאנט(Vanishing Gradient), מה שמאפשר למודול למדוד במהירות ובייעילות (Wu, 2017).



איור 15 : פונקציית ReLU (Wu, 2017).

מלבד ReLU, קיימות פונקציות נוספות כגון softmax, המשמשת בשכבות הפלט לשיווג הסתברותי  $(f(x) \in [0,1])$ .

אחר הפעלת פונקציית האקטיבציה, הפלט מועבר לשכבות הבאות של הרשת, בהן מתבצעים צמצום ממדים (Pooling) והמשך העיבוד לקריאת השיווג הסופי.

#### 4. שכבות ארגנציה (Pooling)

שכבות הארגנציה, או שכבות Pooling, מצמצמות את כמות הנתונים ו לשמור את המידע החיוני מהשכבות הקונבולוציוניות. שכבות אלו פועלות על ידי הקטנת ממד הפלט שהתקבל לאחר הפעלת הפילטרים הקונבולוציוניים, מה שפחית את מספר הפרמטרים בראש ומונע למידת יתר (Overfitting). בנוסף, Pooling מוגבר את עמידות הרשת לשינויים קטנים בתמונה, כגון תזוזות, קלות, שינוי תוארה או עיוותים מינימליים (Wu, 2017).

שכבה Pooling פועל על חלונות קטנים בתמונה, כאשר בכל אזור מתבצע חישוב של ערך ייחיד שמייצג את המידע המרכזי באותו אזור. שתי השיטות הנפוצות ביותר הן Max Pooling ו-Average Pooling.

$$y(i^{l+1}, j^{l+1}, d) = \max_{0 \leq i < H, 0 \leq j < W} X_{i^{l+1} \times H + i, j^{l+1} \times W + j, d}^l$$

שיטה זו מאפשרת שימור של התכונות החזקות ביותר בתמונה וכן נפוצה במודלים של זיהוי תמונה. (Wu, 2017).

לעומת זאת, Average Pooling מחושב באמצעות ממוצע הערכים באזור הנתון, לפי הנוסחה:

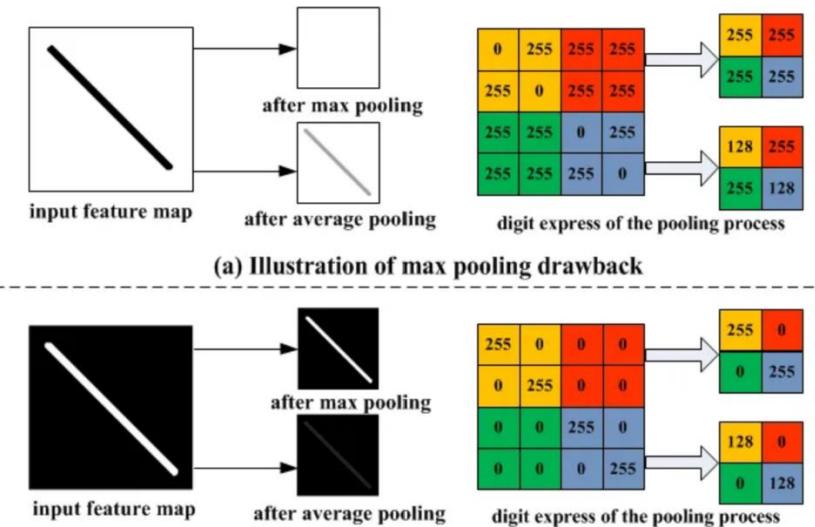
$$y(i^{l+1}, j^{l+1}, d) = \frac{1}{HW} \sum_{0 \leq i < H, 0 \leq j < W} X_{i^{l+1} \times H + i, j^{l+1} \times W + j, d}^l$$

שיטה זו מספקת ייצוג ממוצע של המידע באזור מסוים, אך עלולה לגרום לטשטוש פרטים חשובים, וכן משמשת פחות בקשרים של זיהוי תמונה. (Wu, 2017).

השימוש בגודל חלון ובצעד (Stride) משפייע ישירות על גודל הפלט המתקבל -  $\times H^{l+1} \times W^{l+1}$ . חישוב הגודל נעשה לפי הנוסחה:

$$H^{l+1} = \frac{H^l}{H}, \quad W^{l+1} = \frac{W^l}{W}, \quad D^{l+1} = D^l$$

תהליך זה מפחית את כמות המחשבות, ומאפשר לרשותה ללמידה מיידית רלוונטי מבלי להעמיס על מודל הלמידה. (Wu, 2017).



איור 16 : המראה של הפעלת Max Pooling ו- Average Pooling וההבדל בינהם [18].

חשיבות השימוש ב- Pooling נובעת מהיכולת של השכבה לצמצם את המידע בצורה חכמתית, תוך שימירה על דפוסים עיקריים המופיעים בתמונה. היכולת של Pooling להפחית את ממדיהם הנזונים מקטינה משמעותית את מרכיבות החישוב הנדרשת, מה שמאפשר למודלים עמוקים לפעול במהירות וביעילות רבה יותר (Krizhevsky et al., 2017). בנוסף, לאחר שהתמונה מכילה פרטים רבים שלעיטים אינם קריטיים לזיהוי תבניות מרכזיות, השימוש ב- Pooling מאפשר להתמקד באזורי החשובים ביותר ומנוע הצפה של המודל במידע לא רלוונטי. כמו כן, העובדה שהשכבה מבצעת הפקחת ממדיים משפרת את הכללת המודל (Generalization), ככלומר מאפשרת לו לזהות תבניות חוזרות גם בתמונות חדשות, ללא תלות במיקום המדויק שלהם בתמונה (Wu, 2017).

## 5. שכבת השטחה (Flattening)

לאחר שלב חילוץ המאפיינים באמצעות שכבות הקונволוציה ו- Pooling, הנתונים נשמרים כמבנה תלת-ממדי הכלל גובה (Height), רוחב (Width) ועומק (Depth). לאחר שכבות ה- Fully Connected דורשות קלט חד-ממדי, יש צורך להמיר את מפת התכונות לפורתט של ווקטור חד-ממדי, תהליך המכונה השטחה (Flattening).

אם לדוגמה התקבלה מפת תכונות בגודל  $512 \times 7 \times 7$  שכבת ההשטחה תמיר אותה לווקטור באורך

$$X_{\text{flattened}} \in R^{7 \times 7 \times 512} = R^{25,088}$$

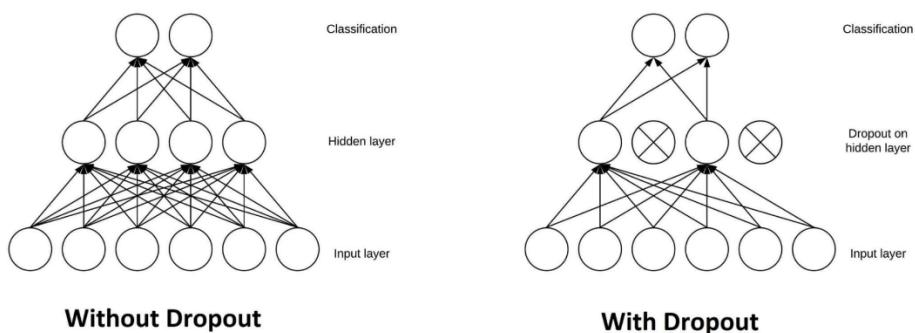
תהליך זה אינו מוסיף פרמטרים חדשים לרשת, אלא מבצע ארגון מחדש של הנתונים כך שניתן יהיה להזין אותם לשכבות Fully Connected, האחראיות על קבלת החלטות הסופיות.

## 6. שכבת Dropout

שכבת Dropout היא טכניקה להפחיתת Overfitting ברשותנו נוירונים עמוקים. במהלך האימון, כל נוירון בשכבות הנסתרות מוושבת בהסתברות של 0.5, ככלומר התפוקה שלו מוגדרת לאפס. נוירונים אלה אינם תורמים להעברת המידע קדימה ואינם משתתפים בחישוב ה- Backpropagation. כתוצאה לכך, בכל שלב של אימון הרשת, היא פועלת על בסיס ארכיטקטורה שונה, אך כל הארכיטקטורות הללו חולקות את אותן משקלים (Krizhevsky et al., 2012).

היתרון המרכזי של Dropout הוא הפחתת התלות של נוירונים מסוימים בנוירונים אחרים, מה שמאפשר את הרשות ללמידה תכונות חזקות יותר שיכולה לפעול בשילוב עם ת-קבוצות שונות של הנוירונים האחרים. בזמן הבדיקה, כל הנוירונים משמשים, אך הפלט שלהם מוכפל ב-0.5, מה ששימוש כאומדן ממוצע גיאומטרי של התוצאות של כל ארכיטקטורות ה-Dropout האפשריות (Krizhevsky et al., 2012).

הטchnique נמצאה כיעילה במיוחד בשכבות Fully Connected והיא שיפרה משמעותית את ביצועי הרשת בתחרות ILSVRC-2012, שם הגיעו לשיעור שגיאה נמוך ב-10.9% בטופ-5 בהשוואה לשיטות אחרות (15.3%) לעומת 26.2% (Krizhevsky et al., 2012).



איור 17 : המחשב השימוש ב [19] Dropout .

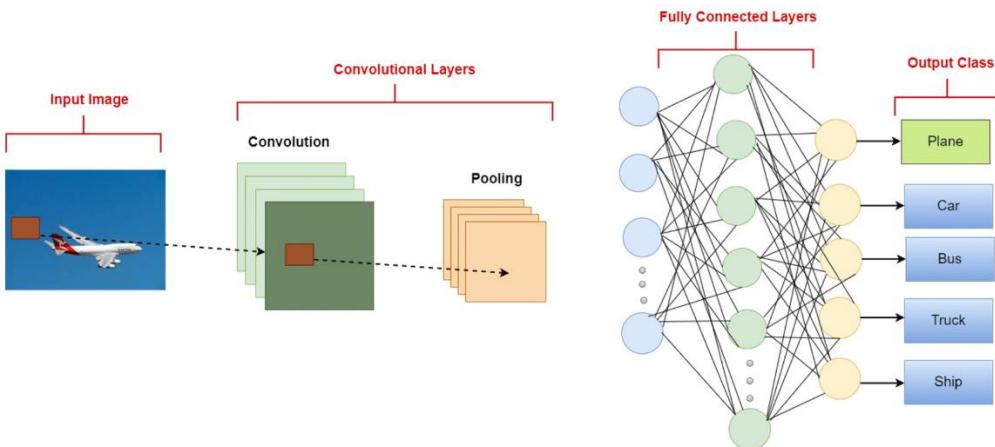
## 7. שכבות Fully Connected

שכבות Fully Connected הן השכבות הסופיות ב CNN, והן משמשות לתהליכי הסיווג הסופי של הנתונים שעברו דרך השכבות הקונבולוציוניות וה Pooling . בשכבות אלו, במקומות להפעיל פילטרים מוקומיים כמו בשכבות הקונבולוציה, כל נוירון מחובר לכל נוירון בשכבה הבאה, בדומה לרשותות נוירונים קלאסיות. מטרתן של שכבות Fully Connected היא לקבל את התכונות שחולצו מהתמונה ולתרגם אותן לסיווג סופי של הרשת, כמו קביעת הרגש המתאים לתמונה.

לאחר שעוברים את שלב ה- Flattening , מתקובל ווקטור חד-ממדי של ערכים מספריים, אשר מהוות את הקלט לשכבה ה- Fully Connected . כל נוירון בשכבה מקבל את כל הנוירונים מהשכבה הקודמת ומבצע את החישוב הבא :

$$y_i = f \left( \sum_{j=1}^N w_{ij} x_j + b_i \right)$$

כאשר  $y_i$  הוא פלט הנוירון בשכבה ה- Fully Connected ,  $x_j$  הוא הערך מהנוירון  $j$  בשכבה הקודמת,  $w_{ij}$  הוא המשקל המחבר בין הנוירון  $i$  לנויירון  $j$  להטיה (Bias) ו-  $f(x)$  היא פונקציית האקטיבציה, כמו ReLU או Softmax .



אייר 18 : התמונה מציגה רשת CNN, ניתן לראות את ההבדל בין קונבולוציה, המבצעת מיפוי מקומי ושימור מרחבי, לבין .(Zafar et al., 2024), שבו כל הנוירונים מחוברים זה לזה ללא תלות במיקום (Fully Connected).

מודלים מודרניים כמו AlexNet עושים שימוש בשלוש שכבות Fully Connected בשלב הסופי של הרשת, כאשר השכבה האחורית מספקת פלט של הסטבריות באמצעות Softmax, המשמשת לחיזוי קטגוריות הסיווג הסופית. עם זאת, בשל המספר הרב של הפרמטרים בשכבות אלו, קיימת סכנה ל-Overfitting, וכן נעשה שימוש בטכניקות כמו Dropout, דבר המחייב את הרשות ללמידה ייצוגית חזקית יותר שאינם תלויים בנוירונים מסוימים (Krizhevsky et al., 2017).

#### 8. שכבת הפלט (Output Layer)

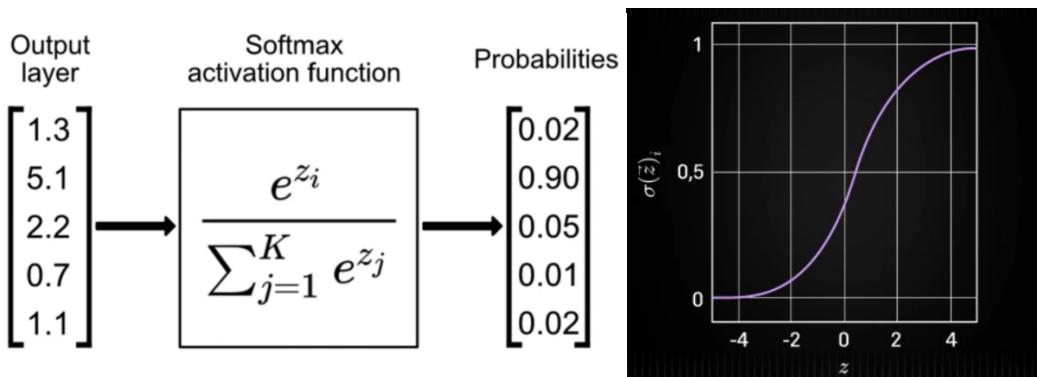
שכבת הפלט היא השלב האחרון ברשת CNN, והיא ממירה את המידע המספרית שהופק משלב הפלט הקודומות לסיווג סופי. השכבה מכילה מספר יחידות פלט השווה למספר הקטגוריות האפשריות, למשל 7 רגשות במודל לזיהוי רגשות. במשימות סיווג מרובות קטגוריות, נעשה שימוש בפונקציית Softmax, אשר ממירה את הפלט להסתברויות כך שסכום יהיה 1 :

$$P(y = k|x) = \frac{e^{z_k}}{\sum_{j=1}^K e^{z_j}}$$

כאשר  $z_k$  הוא הפלט מהנוירון המתאים לקטgorיה  $k$ ,  $K$  הוא מספר הקטגוריות ו( $y = k|x$  מייצג את ההסתברות שהתמונה מסווגת לקטgorיה  $k$ ).(Vasyltsov & Chang, 2021)

הרשת לחברת את כל הנוירונים הקודמים בשכבה הסופית ומבצעת חישוב משוקל על הערכים לאחר חישוב ההסתברויות, הרשת בוחרת את הקטgorיה עם ההסתברות הגבוהה ביותר באמצעות הפעולה :

$$\hat{y} = \arg \max P(y = k|x)$$



איור 19 : פונקציית softmax ודוגמא לפלט הפונקציה [22].

### מימוש CNN ליזהוי רגשות

מימוש מוצלח של רשתות נוירונים קובנולוציוניות (CNN) ליזהוי רגשות מחייב בחירה זהירה של ארכיטקטורת הרשת והיפר-פרמטרים מתאימים. עומק הרשת חייב להיות מאוזן כך שיאפשר למודל ללמידה ייצוגים מורכבים מבלי לגרום ללמידה יתר. רשתות עמוקות מדי נוטות ללמידה יתר, בעיקר, כתוצאה מAGER הנתונים מוגבל, בעוד שרשתות רדודות עלולות לסייע ללמידה חסר (Li & Deng, 2020). מספר הפילטרים בכל שכבה קובנולוציונית משפיע ישירות על יכולת המודל לזהות תבניות מגוונות של הבעות פנים.

בחירת האופטימיזציה המתאימה משפיעה על מהירות ההתקנסות ואיכות הפתרון. אלגוריתמים כמו SGD ו-Adam עם מומנטום מספקים יתרונות שימושיים בניהול קצב הלמידה ובהתאמות משקליות. קצב הלמידה (Learning Rate) מהווה היפר-פרמטר קריטי, כאשר אסטרטגיות כמו Cyclical Learning Rates מאפשרות שיפור ביצועים. גודל-batch המשפיע הן על זמן האימון והן על איכות ההתקנסות. ערכים גדולים מאפשרים חישוב מקבילי עילית אך עלולים להוביל למינימום מקומי גורע, בעוד שערכים קטנים מסייעים בהימנע ממינימום מקומי אך מאריכים את זמן האימון. עדכון המשקלים באמצעות האופטימיזטור Adam מתבצע לפי המשוואה:

$$\theta_t = \theta_{t-1} - a \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon}$$

כאשר  $\theta_t$  מייצג את המשקלים בזמן  $t$ .  $a$  הוא קצב הלמידה.  $\hat{m}_t$  הוא הגרדיאנט הממוצע.  $\hat{v}_t$  הוא הערך הריבועי הממוצע של הגרדיאנטים.  $\epsilon$  הוא מספר קטן למניעת חלוקה באפס (Kingma & Ba, 2014).

יזהוי מוקדם של למידת יתר מתבצע באמצעות ניטור הפער בין ביצועי המודל על סט האימון לסט הווolidציה. אם ביצועי המודל משתפרים על סט האימון אך מתדרדרים על סט הווolidציה, ישנה אינדיקציה ללמידה יתר.

אסטרטגיות להפחחת למידת יתר כוללות Data Augmentation, היוצרת דוגמאות נוספות באמצעות טרנספורמציות על הנתונים המקוריים, ו-Early Stopping, המפסיק את האימון כאשר ביצועי המודל על סט הווolidציה אינם משתפרים יותר. Cross-Validation (K-Fold) מאפשר שימוש עילית יותר בנתונים הזמינים והערכת הכללה מדוקנית יותר על ידי ולידציה של 1 מתוך K קבוצות, והחלפת

הקבוצה המשמשת לוולידציה K פעמים, ומתקבל ממוצע ביצועים המפחית תלוות בחלוקת ספציפית של הנתונים (Li & Deng, 2020).

בחישוב פונקציית ההפסד במשימות סיוג רב-קטגוריות, פונקציית Cross-Entropy היא Categorical Cross-Entropy והיא מוחשבת כך:

$$L_{CE} = - \sum_{i=1}^C y_i \cdot \log(\hat{y}_i)$$

כאשר  $C$  הוא מספר הקטגוריות,  $y_i$  הוא האינדיקטור של הקטgorיה האמיתית,  $\hat{y}_i$  היא ההסתברות שיחסב המודל עבור הקטgorיה  $i$  (Terven et al., 2023). במקרים של חסר איזון בין הקטgorיות, נעשה שימוש ב Focal Loss, המחזקת דוגמאות קשות ומחילה דוגמאות קלות לסיוג:

$$L_{FL} = - \sum_{i=1}^C (1 - \hat{y}_i)^\gamma \cdot y_i \cdot \log(\hat{y}_i)$$

כאשר  $\gamma$  הוא פרמטר המתמקד בדוגמאות קשות (Terven et al., 2023).

הערכת מודל CNN לזיהוי רגשות מבוצעת באמצעות מספר מדדים מעבר לדיווק (Accuracy), כולל F1-Score ו Recall ,Precision

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

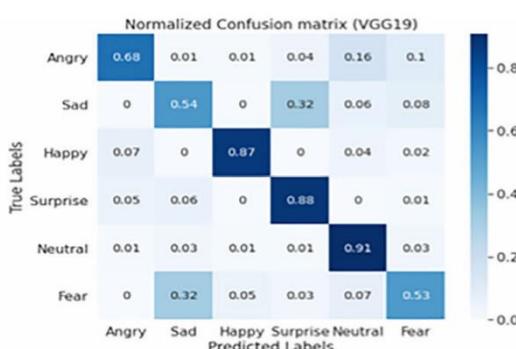
$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$F1 - score = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}$$

כאשר TP הם True Positives, FN הם False Negatives, FP הם False Positives, TN הם True Negatives (Joseph et al., 2024).

שימוש במטריצה בלבול (Confusion Matrix) מאפשר ניתוח דפויי שגיאות והבנת תחומי השיפור של המודל.



איור 20 : דוגמא לתוצאות על Confusion Matrix עם אלגוריתם VGG19 FER-2013 (Gupta, Kumar & Tekchandani, 2023)

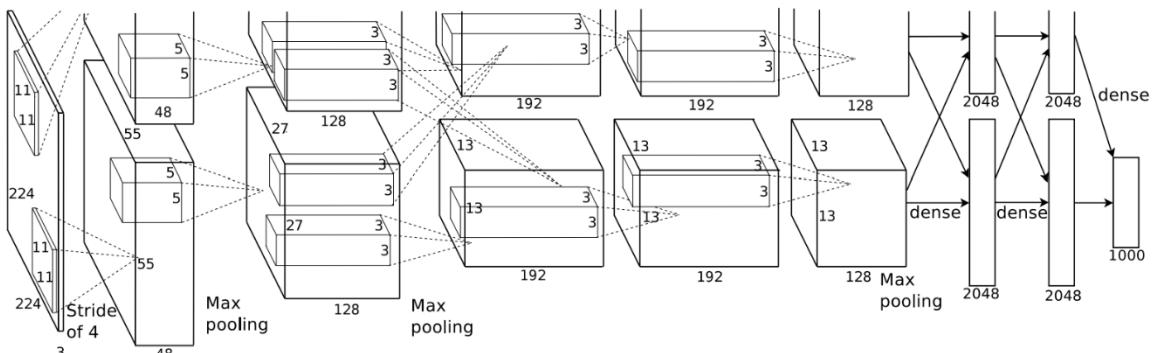
## סקירת מודלים

### AlexNet

רשת AlexNet היא אחת מפראיצות הדרך המשמעותית בתחום הלמידה העמוקה, שהביאה לשיפור ניכר בזיהוי אובייקטים בתמונות. המודל פותח על ידי Krizhevsky, Sutskever ו-Hinton בשנת 2012. הרשת זכתה בתחרות ImageNet Large Scale Visual Recognition Challenge בשנת 2012. הרשת הצינה ביצועים חסרי תקדים והובילה להכרה רחבה ביכולות של רשתות נוירוניות קונבולוציוניות (CNN) במגוון שימושים בתחום הראייה הממוחשב (Krizhevsky et al., 2012).

AlexNet מורכבת משכבות נלמודות, הכוללות חמישה שכבות קונבולוציוניות ושלוש שכבות Fully Connected. המבנה הספציפי של הרשת מפורט להלן:

השכבה הראשונה מקבלת קלט בגודל  $3 \times 224 \times 224$  (RGB). היא כוללת 96 מסננים בגודל  $11 \times 11 \times 3$  עם stride של 4. השכבה השנייה כוללת 256 מסננים בגודל  $5 \times 5 \times 48$  ומבצעת בתחליתה polling וニアרמול. השכבה השלישית, הרביעית והחמישית מחוברות זו לזו ללא שכבות pooling או נירמול. השכבה הששית מורכבת מ-384 מסננים בגודל  $3 \times 3 \times 256$  ומחוברת לשכבה השניה אחרי נירמול ו pooling. השכבה השביעית מכילה 384 מסננים בגודל  $3 \times 3 \times 192$ . השכבה החמישית כוללת 256 מסננים בגודל  $3 \times 3 \times 3$ , ולאחריה מבוצע שלב נוסף של max-pooling עם 4096 נוירונים בשתי השכבות הראשונות. שלוש השכבות הסופיות הן שכבות Fully Connected. השכבה האחורונה 1000 נוירונים, בהתאם למספר הקטגוריות בתחרויות ImageNet. במהלך אימונו הרשת, פلت השכבה האחורונה עבור דרך פונקציית Softmax להפקת התפלגות הסתברויות על פני 1000 הקטגוריות (Krizhevsky et al., 2012).



איור 21 : ארכיטקטורת AlexNet (Krizhevsky et al., 2012)

אחד החידושים הבולטים ביותר ב-AlexNet הוא השימוש ב-ReLU (Rectified Linear Units). במקום פונקציות הפעלה המסורתיות Sigmoid ו-Tanh, נועד על כך, כדי להתמודד עם בעיית למידת יתר (Overfitting), נעשה שימוש בשתי הטכניקות Data Augmentation ו-Dropout.

אחד החידושים המרכזיים של AlexNet היה השימוש בкарטייסי GPU (Graphics Processing Unit) לצורך חישובים מקבילים. בזמן אימונו המודול, המפתחים חילקו את החישובים בין שני כרטיסים מסוג NVIDIA GTX 580, כך שכל חלק שונה של הרשת, תוך מינימום תקשורת ביניהם.

גישה זו אפשרה אימון רשת בקנה מידה גדול באופן שלא היה מעשי עד אותה תקופה (Krizhevsky et al., 2012).

AlexNet השיגה שיעור טעות של 15.3% בתחרות ILSVRC-2012, שיפור משמעותי ביחס ל-26.2% שהושגו על ידי המודל במקום השני. עובדה זו הפכה אותה לאבן דרך חשובה בהתפתחות הלמידה העמוקה, שכן היא הוכיחה את יכולת רשות נוירונים עמוקים לתמודד עם שימוש סיווג מורכבות (Krizhevsky et al., 2012).

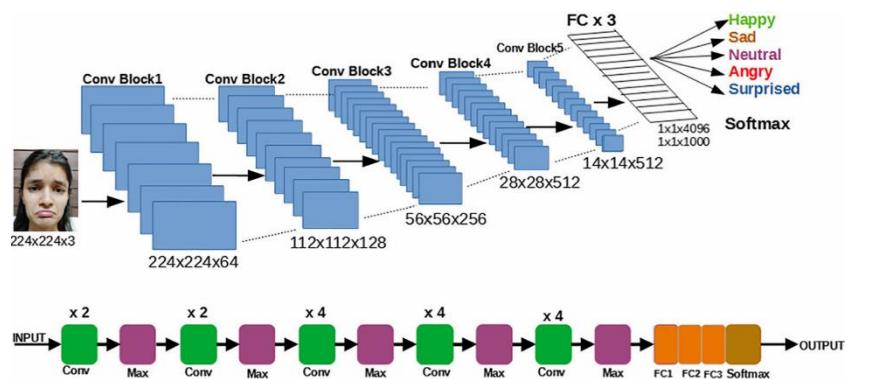
### VGG-19

רשת VGG-19 היא אחת מהרשתות הנוירוניות הקונבולוציוניות (CNN) המשמשות ביוטר בתחום ראייה ממוחשבת, והיא פותחה על ידי Simonyan & Zisserman כחלק מחקרים באוניברסיטת אוקספורד. הארכיטקטורה של VGG-19, הידועה בפשוטה ובעומקה, נועדה לשפר את עיצוב הסיווג על ידי שימוש במסננים קטנים והעמקת הרשת, תוך שמירה על מבנה היררכי ברור של שכבות (Gupta et al., 2022).

רשת VGG-19 מורכבת מ-19 שכבות למידה, הכוללות 16 שכבות קונבולוציה עם מסננים בגודל  $3 \times 3$  וחמש שכבות Fully Connected, Max-Pooling, ושלוש שכבות Max-Pooling. המבנה הספציפי של הרשת הוא כדלקמן :

קלט הרשת מתקיים בתמונה  $3 \times 224 \times 224$  (RGB).  
שתי שכבות קונבולוציה בגודל  $3 \times 3$  עם 64 מסננים, ולאחריהן שכבה Max-Pooling בגודל  $2 \times 2$ .  
שתי שכבות קונבולוציה נוספת בגודל  $3 \times 3$  עם 128 מסננים, ולאחריהן שכבה Max-Pooling בגודל  $2 \times 2$ .  
ארבע שכבות קונבולוציה בגודל  $3 \times 3$  עם 256 מסננים, ולאחריהן שכבה Max-Pooling בגודל  $2 \times 2$ .  
ארבע שכבות קונבולוציה בגודל  $3 \times 3$  עם 512 מסננים, ולאחריהן שכבה Max-Pooling בגודל  $2 \times 2$ .  
ארבע שכבות קונבולוציה בגודל  $3 \times 3$  עם 512 מסננים, ולאחריהן שכבה Max-Pooling בגודל  $2 \times 2$ .  
שלוש שכבות Fully Connected - שתי שכבות עם 4096 נוירונים, ושכבה שלישית עם 1000 נוירונים, התואמת למספר הקטגוריות של ImageNet.

פלט הרשת מתקיים דרך פונקציית Softmax, אשר מפיקה התפלגות הסתברויות על פני 1000 הקטגוריות האפשריות (Gupta et al., 2022).



איור 22 : ארכיטקטורת VGG-19 (Gupta et al., 2022)

ייחודה של VGG-19 טמון בשימוש במסננים קטנים בגודל  $3 \times 3$  בכל שכבות הקונבולוציה. גישה זו מאפשרת לרשת למדוד מאפיינים מורכבים יותר על ידי שילוב רציף של מסננים קטנים, תוך הקטנת מספר הפרמטרים לעומת שימוש במסננים גדולים יותר (Gupta et al., 2022).

כדי למנוע במידה יתר (Overfitting) ולשפר את הביצועים, הרשת עשו שימוש בטכניקות Dropout ו- Regularization (Gupta et al., 2022) Batch Normalization.

רשת VGG-19 מציגה מספר יתרונות משמעותיים שהופכים אותה לכלי מרכזי בראיה ממוחשבת. אחד מה יתרונות הבולטים ביותר של הארכיטקטורה הוא מבנה השכבות הסדור והעמוק, אשר מאפשר למודל למדוד ייצוגים מורכבים יותר בהשוואה לרשומות קודמות. גישה זו מספקת יתרון משמעותי כאשר מדובר במשימות כמו סיוג תמונות וזיהוי אובייקטים, שכן היא מאפשרת לרשת לזהות תבניות שונות תוך שימירה על היררכיה המאפיינית. לאחר שהרשת אומנה מראש על מערך הנתונים הגדול ImageNet, ניתן לבצע התאמת מהירה של המודל למשימות חדשות תוך חיסכון משמעותי בזמן חישוב ומשאבי אימון.

עם זאת, לצד יתרונותיה, קיימות מגבלות המהוות חסרונות מרכזיים בשימוש ב-VGG-19. אחד האתגרים העיקריים של הרשת הוא מספר הפרמטרים הגבוה, שעומד על כ-144 מיליון פרמטרים, מה שהופך אותה לקרה מבחינה חישובית ודורשת משאבי חומרה מתקדמים, כגון יחידות עיבוד גרפיות (GPU) חזקות וזריכתן רב. בנוסף, בשל עומקה הרב, הרשת אינה יעילה מבחינת זמן חישוב, שכן כל שכבת קונבולוציה מסוימת שכבת חישוב נוספת, מה שוביל לעלייה משמעותית בזמן האימון והשימוש (Gupta et al., 2022).

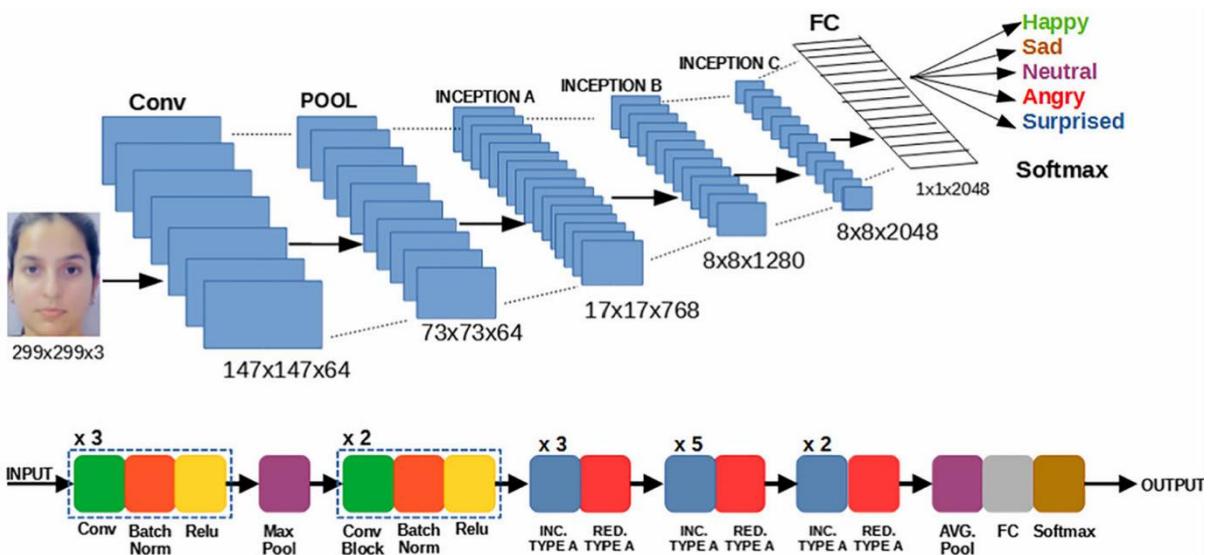
### Inception-V3

רשת Inception-V3 היא ארכיטקטורה מתقدמת של רשתות ניורוניים קונבולוציוניות, אשר פותחה כהמשך לשיפורים שנעשו ב-Inception-V1 (GoogLeNet). הרשת נועדה להתמודד עם אתגרים בתחום סיוג התמונות, תוך שיפור הדיוק ושמירה עליעילות חישובית גבוהה. אחד המאפיינים המרכזיים של Inception-V3 הוא השימוש בגישות חישובית שטטרן להפחית את מספר הפרמטרים הנדרשים למודל, ובכך להפוך אותו לאפקטיבי יותר הן מבחינת הביצועים והן מבחינת השימוש במשאבי עיבוד. במחקר שבו נבדקה רשת זו בהקשר של זיהוי רגשות על סמך הבעות פנים, היא נבחנה לצד מודלים אחרים, במטרה להעריך את מידת הדיוק שלה ואת יכולתה להבחן בין רגשות שונים בסביבה דינמית של במידה מקוונת (Gupta et al., 2022).

אחד המאפיינים הייחודיים של רשת זו הוא המבנה ההיררכי של שכבות ה-*pooling*, המאפשר שילוב של מסננים גדולים שונים בכל שכבה, כך שהרשת מסוגלת למדוד ייצוגים מגוונים בו-זמנית. הרשת מקבלת קלט תמונה בגודל  $224 \times 224 \times 3$ , אשר עוברת עיבוד מקדים באמצעות שכבות קונבולוציה ראשונית. לאחר שלבי העיבוד המקדמים, הרשת מחלוקת למספר בלוקים מרכזיים הנקראים Pooling, שכבות Inception. בלוקים אלה מספקים לרשת את יכולת לזהות תבניות ברמות שונות של הפשטה, תוך שימוש במבנה חישובי חסכוני.

אחד השיפורים המשמעותיים ב-Inception-V3 הוא השימוש בפקטוריזציה של קובולוציות, אשר מאפשרת להמיר מסננים גדולים למספר מסננים קטנים יותר, וכך להפחית את כמות החישובים הנדרשת. במקרה להשתמש במסנן  $5 \times 5$ , נעשה שימוש בשני מסננים עוקבים בגודל  $3 \times 3$ , מה שמאפשר הפחיתת מספר הפרמטרים ושיפור יעילות החישוב. נוסף לכך, הרשות כוללת בלוקים ייחודיים המכונים Reduction Blocks, שתפקידם לצמצם את מידת הנתונים ולאפשר חישוב מהיר יותר מבלי לאבד מידע מהותי. טכניקות נוספות, כגון Dropout ו-Batch Normalization, מסייעות בייצוב הרשות ומונעות למידת יתר. שיטות אלו, לצד Label Smoothing, מסייעות לשיפור דיוק הסיווג באמצעות הפחיתת התלות בדוגמאות מסוימות במהלך האימון (Gupta et al., 2022).

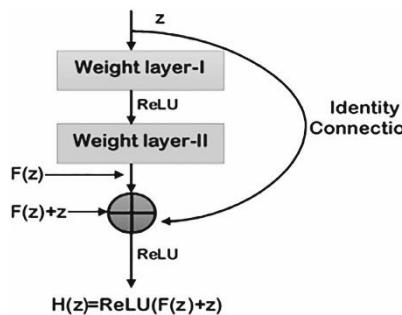
במחקר שבו נבחנה רשת Inception-V3-Plus במשימות של זיהוי רגשות, היא הושווה למודלים אחרים ונמצאה יעילה במיוחד. תוצאות הניסוי הראו כי הרשות השיגה דיוק של 80.76% במאגר CK+. Inception-V3 יתרו על שיעור מושגים מבחן האיזון בין ביצועים לדרישות חישוביות. המבנה שלה מאפשר חישוב מהיר יותר לעומת רשתות עמוקות יותר כמו ResNet-50, מה שהופך אותה לאופטימלית עבור מערכות הדורשות זמן תגובה מהיר, במיוחד כאשר משאבי החומרה מוגבלים. השימוש בבלוקים של Inception מאפשר לרשת לקלוט מידע מרמות שונות של מרכיבות באופן מקבילי, מה שמאפשר לרשת לזהות קשרים בין פרטיים קטנים וגדולים באותו תמונה. שיטה זו מקנה לרשת גמישות גבוהה יותר ומונעת ממנה להתמקדש יתר על המידה בפרטיים קטנים בלבד או במקרים רחבים בלבד. בנוסף, היכולת לבצע עיבוד במספר רמות בו-זמנית מפחיתה את הסיכון למידת יתר, משום שהרשת אינה נדרשת להתחאים את עצמה באופן הדוק כדי לדוגמאות אימון מסוימות. רשת Inception-V3 מושיכת להיות אחת הרשות היעילות ביותר לסיווג תמונות הودות לאיזון שהיא מציעה בין דיוק לעילوت חישובית. בעוד שהדיק שלה במשימות זיהוי רגשות היה מעט נמוך בהשוואה ל-ResNet-50, היא עדין מציעה יתרונות משמעותיים מבחן חיסכון במשאבי חישוב ושיפור הביצועים בזמן אמת. שילוב של טכניקות מתקדמות להפחית עומס חישובי תוך שימירה על ביצועים גבוהים הופך את Inception-V3 למודול מרכזי בתחום זיהוי פנים, סיווג תמונות, ויישומי ראייה ממוחשבת נוספים (Gupta et al., 2022).



.(Gupta et al., 2022) Inception-V3 אירור 23 : ארכיטקטורת

## ResNet-50

רשת ResNet-50 היא אחת הארכיטקטורות המשפיעות ביותר בתחום הראייה הממוחשבת, והיא מבוססת על עקרון הרשות השיווריות (Residual Networks). ארכיטקטורה זו פותחה במטרה להתגבר על בעיית דעיכת הגרדיינט (Vanishing Gradient Problem), אשר מהוות אתגר מרכז'י באימון רשתות עמוקות מאוד. ביגוד לרשותות מסורתיות, ResNet-50 עושה שימוש בקשרים שיווריים (Skip Connections), המאפשרים למידע לירום דרך שכבות הרשת באופן ישיר, ובכך נמנע איבוד מידע לאורך עומק הרשת (Gupta et al., 2022).



איור 24 : מבנה בLOC שיורי ברשת ResNet-50  
(Gupta et al., 2022)

רשת ResNet-50 כמודל עמוק הכולל 50 שכבות נלמדות. היא מתחילה בקבלת תמונה בגודל  $3 \times 224 \times 224$ , אשר עברת עיבוד מקדים דרך שכבת קונבולוציה בגודל  $7 \times 7$  עם 64 מסננים ו- Strides בגודל 2. שלב זה מבצע חישוב ראשוני ומיצמץ את ממדיה התמונה. לאחר מכן, מתבצע Max Pooling בגודל  $3 \times 3$  עם Stride של 2 כדי להפחית את ממדיה הנתונים לפני הכניסה לבlocs שיווריים, ולשפר את היעילות החישובית של המודל (Gupta et al., 2022).

לאחר שלב העיבוד הראשוני, הרשת מרכיבת ארבעה שלבים מרכז'יים, כאשר בכל שלב מושלבים blocים שיווריים (Residual Blocks) הקיימים שלוש שכבות קונבולוציה בגודל  $1 \times 1$ ,  $3 \times 3$  ו-  $1 \times 1$ . שכבת ה- $1 \times 1$  הראשונה משמשת להפחיתת ממדיה הנתונים לפני החלת המסלנים הגדולים יותר. שכבת ה- $3 \times 3$  המרכזית אחראית לממדית הבניות עמוקות יותר. שכבת ה- $1 \times 1$  האחורה מחזירה את הממדים המקוריים של המידע שהוזן לבLOCK. גישה זו מאפשרת לרשת לשמור על מספר נמוך של פרמטרים תוך שיפור הדיוק, וזאת באמצעות ניהול חכם של הפחיתת והחזרת הממדים בתוך כל BLOCK שיוורי.

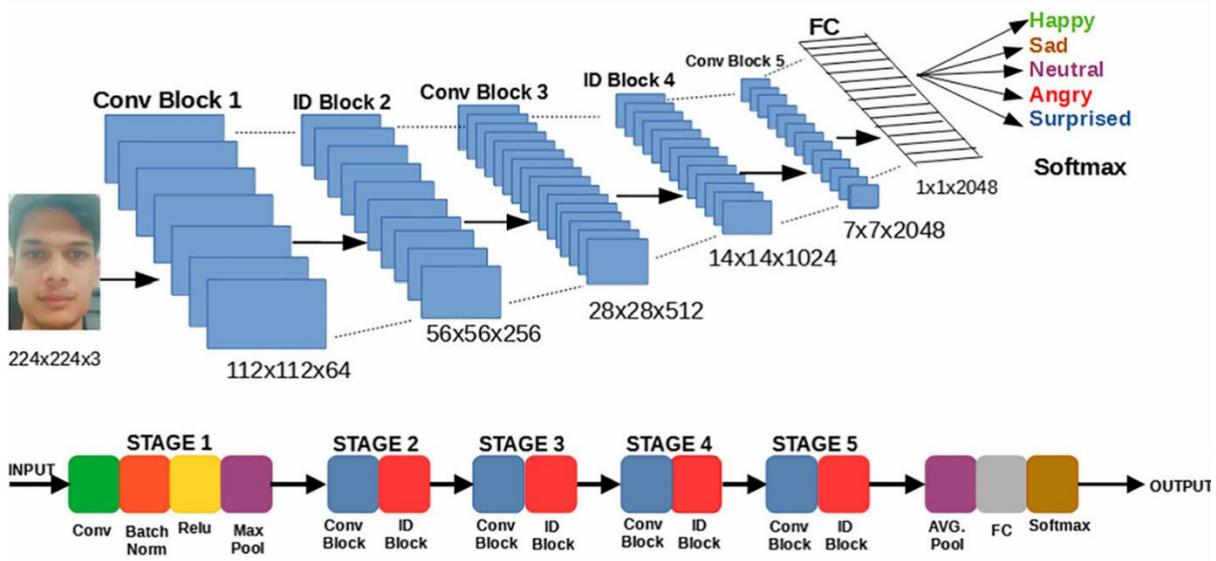
השלב השני מקבל את הקלט מה-Max Pooling הראשוני ומעביר אותו דרך שלושה blocים שיווריים, כאשר בכל BLOCK יש שלוש שכבות קונבולוציה עם 64, 64 ו- 256 מסננים בהתאם.

השלב השלישי מכיל ארבעה blocים שיווריים, כאשר כל BLOCK כולל שלוש שכבות קונבולוציה עם 128 ו- 512 מסננים בהתאם.

השלב הרביעי כולל שישה blocים שיווריים, כאשר בכל BLOCK יש שלוש שכבות קונבולוציה עם 256 ו- 1024 מסננים בהתאם.

השלב החמישי מרכיב שלושה blocים שיווריים, בהם כל BLOCK מכיל שלוש שכבות קונבולוציה עם 512, 512 ו- 2048 מסננים בהתאם.

בשלב האחרון, המידע עובר דרך השכבות השיווריות עבור דרך שכבת Average Pooling גלובלית, ולאחר מכן מועבר לשכבת Fully Connected וsoftmax, אשר מספקת את הפלט קטגוריה הסופית של הסיווג (Gupta et al., 2022).



.איור 25 : ארכיטקטורת ResNet-50 (Gupta et al., 2022)

מבנה ה- Skip Connections (קישורים שיוריים) הוא אחד המאפיינים החדשניים ביותר של ResNet-50, ומאפשר למודל לשמור מידע קריטי מהשכבות הראשונות גם לאחר הרשות מעמיקה. במקום להעביר את הקלט ישירות דרך השכבות הקונבולוציוניות, הרשות מוסיפה את הקלט המקורי לפלט של הבלוק הקונבולוציוני, וכך מונעת איבוד של מידע חשוב בתהליכי הלמידה. תופעת דעיכת הגראדיאנט עלולה לגרום לכך שהמשקלים בשכבות העמוקות יתעדכנו בצורה מינימלית, מה שモוביל לכך שהמודל יתקשה ללמידהTABENIOTOT. השימוש בקישורים שיוריים מאפשר למודל לעקוף שכבות מסוימות ולשמור על זרימת גראדיאנט יציבה, ובכך מבטיח שתתוכנות הרלוונטיות יŁIMDO בצורה יעילה יותר. גישה זו מאפשרת לרשותות עמוקות בהרבה להמשיך ולהתאמן מבליל פגוע בדיקוק או ביציבותו של המודל (Gupta et al., 2022).

רשת ResNet-50 הציגה ביצועים מצוינים בהשוואה לרשותות אחרות במשימות של זיהוי ורגש זיהוי פנים. במחקר אשר השווה בין ResNet-50, VGG19 ו-Inception-V3, נמצא כי ResNet-50 השיגה את הדיקוק הגבוה ביותר, עם תוצאה של 83.53% בסיווג רגשות בזמן אמת במאגר CK+. יתרונה של ResNet-50 ה证实ה בעיקר בזכות יכולת שלה להתמודד עם וריאציות שונות בנתוני הקלט, מה שהוביל להכללה טובה יותר לעומת המודלים האחרים. עם זאת, המרכיבות החישובית של הרשות גבוהה יותר, וזמן האימון שלה ארוך יחסית. מגבלה זו יכולה להיות חיסרון כאשר יש צורך בפתרון בעיה מבחן חישובית, במיוחד במערכות בזמן אמת או במכשירים בעלי יכולות עיבוד מוגבלות (Gupta et al., 2022).

#### ג. אופטימיזציה

תהליך האופטימיזציה מלא תפקיד מركזי בשיפור ביצועי מודלים של למידה عمוקה, במיוחד כאשר מדובר במערכות מורכבות כמו רשותות נירונים קונבולוציוניות ליזיהו רגשות. אופטימיזציה יעילה מאפשרת למודלים למדוד ב邏輯ות גבוהה הרבה יותר, להפחית את שיעור השגיאות ולשפר את יכולת הכללה שלהם על נתונים חדשים. במערכות ליזיהו רגשות פנים, התאמת מדויקת של היפר-פרמטרים, כגון קצב הלמידה, משקלים ראשוניים, ופונקציות האקטיבציה, יכולה להשפיע באופן משמעותי על הדיקוק והביצועים הכלולים.

אחד האתגרים המרכזיים באופטימיזציה של רשותות נירונים עמוקות הוא איזון בין חקירה (Exploration) של מרחב הפתרונות לבין ניצול (Exploitation) של פתרונות מבטחים. מצד אחד, יש צורך לבדוק טוח רחב של פרמטרים כדי לאטר את הקונפיגורציה האופטימלית, אך מצד שני, יש למנוע התכונות מוקדמות לפתרון שאינו גלובלי. במטרה להתמודד עם אתגרים אלה, נעשה שימוש באלגוריתמים של אופטימיזציה מטה-יריסטיים, המדמים תהליכי טבעיות כמו אבולוציה, חישוב מבוסס אוכלוסייה והתנהוגיות בעלי חיים.

#### Walrus Optimization Algorithm (WOA)

במודל NS-EWDL-BFSN, המיועד ליזיהו רגשות פנים, נעשה שימוש באלגוריתם Walrus Optimization Algorithm (WOA) לצורך התאמת אופטימלית של היפר-פרמטרים בראשת ה-EK-ResNet50. אלגוריתם זה, המבוסס על דפוסי התנהוגות של סוביים בטבע, נבחר בשל יכולתו לשלב בצוות דינמית חקירה רחבה של מרחב הפתרונות עם עדכון מדויק של פרמטרים קריטיים. האלגוריתם פועל תוך שימוש במנגנוני חישוב והסתגלות, המחקים את אסטרטגיות הרבייה, הנדידה וההתזונה של סוביים, ומספקים דרך חדשה ויעילה למצום ביצועי המודל (Joseph et al., 2024).

אלגוריתם WOA עושה שימוש באיתוטי סכנה ובתיוחות כדי להגיד את אופן החישוב של הפתרונות. בשלב החקירה המוקדם, העדר נע לעבר אזור חדש במרחב הפתרונות אם מתקבלים איתוטי סכנה מסוימים. לעומת זאת, כאשר העדר מתיצב באזור מסוים, האלגוריתם נכנס לשלב הניצול, שבו איתוטי בתיוחות משמשים לקביעת אסטרטגיות החישוב המקומי. במהלך שלב זה, האלגוריתם מחקה את דפוסי התנהוגות של סוביים בעת חישוב מזון או במנוחה, כאשר הפרטים בעדר מתקשרים זה עם זה במטרה ליעיל את תהליך ההסתגלות ולשפר את מיקום הפתרונות במרחב החישוב (Joseph et al., 2024).

האלגוריתם פועל על בסיס עדכון פרמטרים הדרgti, תוך שימוש בשני פתרונות מובילים כדי לשפר את איכות החישוב ולמנוע התכונות מוקדמות לאופטימום מקומי. בכל שלב, המיקום החדש של פרט באוכלוסייה מחושב כממוצע של שני ערכי יעד, אשר נגזרים מהפתרון הטוב ביותר ומהפתרון השני הטוב ביותר שנמצאו עד כה. החישוב מתבצע באמצעות הנוסחה:

$$Y_{j,k}^{u+1} = \frac{Y_1 + Y_2}{2}$$

כאשר הערכים  $Y_1$  ו-  $Y_2$  מייצגים שני משקלים עיקריים המשפיעים על התנועה של הפתרון הנוכחי. החישוב של  $Y_1$  מתבסס על הפתרון הטוב ביותר שנמצא באיטרציה הנוכחיית, כך שהעדכון נעשה לפי הנוסחה:

$$Y_1 = Y_{best}^u - b_1 \times c_1 \times |Y_{best}^u - Y_{j,k}^u|$$

כאשר  $Y_{best}^u$  מסמן את המיקום של הפתרון הטוב ביותר ביחס ל- $Y_{j,k}^u$  מייצג את המרחק בין הפתרון הנוכחי לבין הפתרון האיךוטי ביותר שנמצא עד כה. הפרמטרים  $b_1$  ו-  $c_1$  משמשים לבקרה קצוב העדכון ולוויות מידת התנועה של הפתרון החדש לכיוון פתרון זה. בנוסף, מתבצע עדכון נוסף המבוסס על הפתרון השני בטיבו, המוחש באמצעות:

$$Y_2 = Y_{second}^u - b_2 \times c_2 \times |Y_{second}^u - Y_{j,k}^u|$$

כאשר  $Y_{second}^u$  הוא המיקום של הפתרון השני הטוב ביותר שנמצא, והמרחב בין הפתרון הנוכחי לפתרון זה משמש כגורם נוסף בקביעת הכיוון החדש של החיפוש. מקדמי הבקרה  $b_2$  ו-  $c_2$  מסייעים בהתאמת המהירות והתנועة עבור פתרון זה, באופן דומה לחישוב עבור  $Y_1$ .  
המקדים  $b$  ו-  $c$  קבועים בצורה דינמית כדי לאזן בין חקירה רחבה של המרחב לבין התמקדויות באזוריים מבטיחים. חישוב  $b$  מתבצע באמצעות:

$$b = \chi \times s_1 - \chi$$

כאשר  $\chi$  הוא מקדם דינמי ו-  $s_1$  הוא ערך אקראי הנבחר בטוווח [0,1]. במקביל, המקדים  $c_1$  ו-  $c_2$  מחושבים לפי:

$$c = \tan(\varphi)$$

כאשר  $\varphi$  הוא פרמטר הנבחר בטוווח [0, π], אשר שולט בעוצמת ההתקראבות של הפתרון הנוכחי לעבר הפתרון האופטימלי. שימוש בפרמטרים אלו מאפשר למודל לשמור על איזון בין חקר מרחב הפתרונות לבין שיפור הדרגתית של הפתרון הנוכחי, תוך התאמה מתמדת לתנאי האופטימיזציה (Joseph et al., 2024).

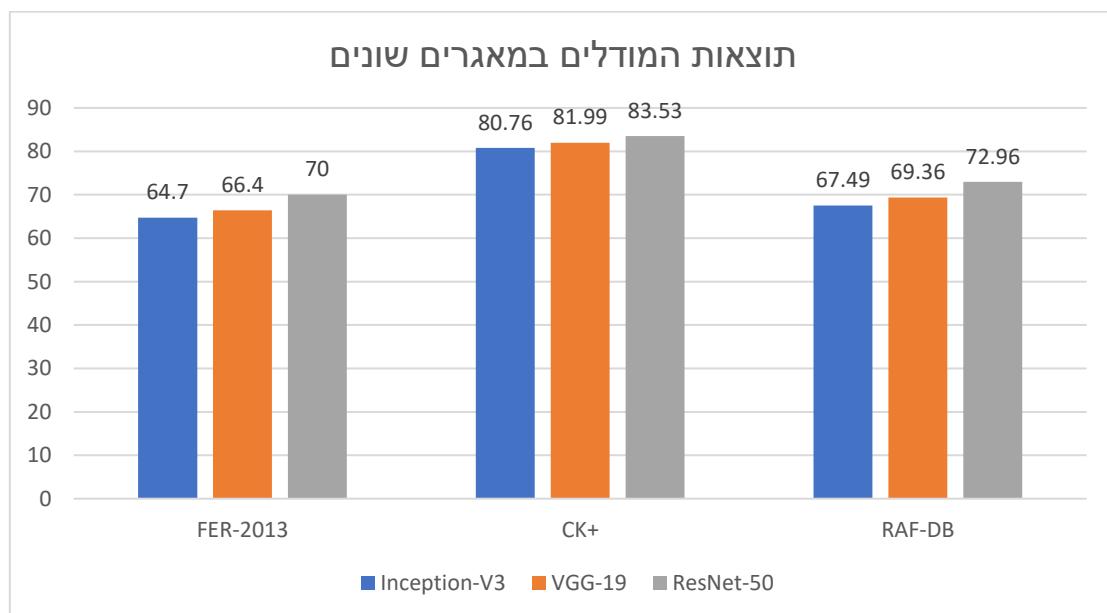
בשילוב לשיטות אופטימיזציה אחרות, כגון (GA) Genetic Algorithm או Particle Swarm Optimization (PSO), נמצא כי WOA מציג יתרון משמעותי ביציבות ובਮהירות ההתכנסות לפתרון אופטימלי. זאת הודות לשימוש במודל התקשורתי החברתי של סוסיים, המאפשר לבצע התאמות עדינות לפרמטרים לאורך שלבי החיפוש והኒצל. כמו כן, האלגוריתם הראה עמידות גבוהה יותר לביעות של התכנסות מוקדמת והטייצבות על פתרונות שאינם אופטימליים, דבר שאופיין לשיטות חיפוש מבוססות אוכלוסייה (Joseph et al., 2024).

## **פרק 5 - מסקנות ודיון**

המחקר הנוכחי סקר את השימוש ברשתות נוירונים קובולוציוניות (CNN) ליזיהו רגשות על בסיס תמונות פנים. בין המודלים שנסקרו נמצאים AlexNet, Inception-V3, VGG-19, ResNet-50 ו-Walrus Optimization Algorithm (WOA), אשר יושם לשיפור היפר-פרמטרים ברשתות הלמידה העמוקה.

השוואה ביצועי המודלים במערך הנתונים FER-2013 הראתה הבדלים בדיק הסיוג. לפי Gupta et al. (2023), רשת ה-ResNet-50 השיגה את הביצועים הגבוהים ביותר, עם דיק של 70%, בהשוואה ל-66.4% residual Inception-V3 ו-64.7% VGG-19. נתוני אלו מצביעים על כך שמנגנון ה-connections residual, המאפשר זרימת מידע ישירה בין שכבות מרוחקות, מסייע בהתמודדות עם בעית דעיכת הגרדיאנט ומשפר את יכולת הכללת המודל. בנוסף, רשתות ה-ResNet לרוב דורשות פחות עיבוד מקדים של הנתונים בשל התאמות המבניות שלהן, מה שעשוי להסביר את יתרונו (Gupta et al., 2023).

במערך הנתונים CK+, אשר מכיל תמונות שנאספו בתנאים מבוקרים, הוצגו תוצאות דומות. ResNet-50 השיגה דיק של 83.53%, כאשר ביצועה היו טובים משמעותית מלה של המודלים האחרים שנבדקו, מה שמעיד על יכולתה לזהות תבניות גס במצבים של שונות מבוקרת (Gupta et al., 2023). לעומת זאת, כאשר נבדקו המודלים על מערך הנתונים RAF-DB, אשר מכיל תמונות שנאספו מסביבות בלתי מבוקרות, חלה ירידת כוללת בדיק הסיוג. עם זאת, גם כאן ResNet-50 הציגו יתרון ייחסי עם דיק של 72.96%, בהשוואה ל-69.36% VGG-19 ו-67.49% Inception-V3. עובדה זו מחזקת את ההשערה כי מבנה ה-residual connections מאפשר למודל לשמור ביצועים טובים יותר בתנאי סביבה מגוונים (Gupta et al., 2023).



איור 26 : ההשוואה בין מודלים במאגרים שונים .(Gupta et al., 2022)

אלגוריתם האופטימיזציה WOA יושם לשיפור ביצועי רשת ResNet50, כחלק ממודל-EWDL-BFSN, שהותאמת לזיהוי רגשות. השימוש ב-WOA מאפשר אופטימיזציה של היפר-פרמטרים והפחיתה שיעור ה-false positives, תוך השגת דיוק של 99.37% CK+ ו- 99.25% FER-2013. תוצאות אלו מדגימות את יכולת של אלגוריתם האופטימיזציה לשפר את ביצועי רשותות במידה עמוקה על ידי כיוון דינמי של הפרמטרים הクリיטיים, אך הם מושפעות גם ממודל EWDL-BFSN המתואר (Joseph et al., 2024).

בשילובת בין המודלים, ניתן לראות כי ResNet-50 מתאפיינת ביכולת הכללה משופרת על פני מערכי נתונים שונים, בעיקר בשל המבנה הייחודי שלה שמנע דעיכת גרדיאנט ומאפשר הקשר יציבה גם ברשותות עמוקה. VGG-19, על אף השימוש במספר רב של שכבות קובולוציה, מתקשה להגיע לרמת הדיק זומות בשל מבנה ליניארי יותר, המוביל לרגישות לשינויים קטנים בקלט. Inception-V3, על אף השימוש באררכיטקטורת inception blocks, מתקשה גם הוא לעמוד בביטחון ResNet-50 בערך כאשר מדובר בסביבות לא מבוקרות, מה שמרמז על מגבלות מסוימות ביכולת הכללה של המודל (Gupta et al., 2023).

לסיכום, השוואת ביצועי המודלים מראה כי ResNet-50 היא הרשת היעילה ביותר לזיהוי רגשות פנים במגוון רחב של תרחישים. האופטימיזציה שבוצעה באמצעות WOA סייעה להגדיל את הדיוק ולהפחית את שיעור השגיאות, תוך שיפור משמעותית בביטחון המודל. עם זאת, קיים צורך בהמשך מחקר שיבחן שיטות נוספות לשיפור ההכללה של המודלים, במיוחד במקרים של וריאציות גבוהות בתנאי התאורה והבעת הרגש.

## פרק 6 - סיכום אישי

במהלך כתיבת העבודה, עסكتי בIMPLEMENTATION של נוירונים קונבולוציוניים ליזיהוי רגשות פנים, תוך חקירה מעמיקה של ארכיטקטורות שונות, עיבוד נתונים, השוואת鄙 ציעום ושימוש בטכניקות אופטימיזציה מתקדמות. העבודה העניקה לי הזרנות לבחון את האתגרים הכרוכים ביישום מודלים של מידת מכונה, הן בהיבטים טכניים והן בהשלכות האתיות של טכנולוגיות ליזיהוי רגשות. מעבר לניתוח והשווות鄙 ציעום, התחרד לי המורכבות הטכנולוגית הקיימת, הכוללת מגבלות חישוביות, תלות נתונים גדולים ושאלות הנוגעות ליכולת ההכללה של המודלים.

אחד ההיבטים המשמעותיים בפרויקט היה ההבנה כיצד שלבי הכתנת הנתונים והשימוש המקדים משפיעים ישירות על鄙 ציעום המודלים. ההתנסות בטכניקות כמו נורמליזציה, הסרת רעשים ו-Data Augmentation מהבינה עד כמה ניהול נתונים טרם שלב האימון הוא קריטי להצלחת המודל. כמו כן, למדתי לעומק כיצד בחירת ארכיטקטורה מסוימת משפיעה על鄙 ציעום הרשות, והבנתי את היתרונות והחסרונות של גישות שונות בתכנון מבנה הרשות. שימושו של AOA כסטרטגיית אופטימיזציה אפשר לי לבחון כיצד יכולותם של היפר-פרמטרים משפר את鄙 ציעום המודלים, מיעיל את תהליכי האימון ומפחית את זמני החישוב.

מעבר לכתיבת הסמינר, אני משתמש גם בסדנה במדעי הנתונים, שבה בחרתי למשתמש בפועל מודל ליזיהוי רגשות באמצעות במידה עמוקה. הבחירה בIMPLEMENTATION זה נבעה באופן טבעי מתוך הידע שרכשתי במהלך העבודה על הסמינר. ההבנה התיאורטית המعمיקה שנבנתה תוך כדי המחקר תאפשר לי בעת לישום את המודל באופן מושכל יותר, תוך בחירה נכונה של ארכיטקטורה, עיבוד נתונים מתאימים והטאמת פרמטרים מיטבית. העבודה על הסמינר סיימה לי כלים להבנה ביקורתית של המודל, כך שבשלב המימוש יוכל לשים דגש על שיפור דיקוק המערכת, מזעור למידת יתר והתמודדות עם מגבלות חישוביות.

פרויקט זה לא רק חיזק את הבנתי בתחום למידת המכונה, אלא גם המשיך את הקשרים ההדוקים בין מחקר תיאורי ליישום מעשי. הוא סייף לי הזרנות לחקר לעומק את התחום של ליזיהוי רגשות, תוך בחינת האתגרים הטכנולוגיים והאתיים הכרוכים בכך, והעניק לי בסיס מוצק להמשך עבודה בתחום זה, הן בהיבט האקדמי והן בפרויקטים יישומיים שאבצע בעתיד.

## **פרק 7 - רשימת מקורות**

1. Pereira, R., Mendes, C., Ribeiro, J., Ribeiro, R., Miragaia, R., Rodrigues, N., ... & Pereira, A. (2024). Systematic Review of Emotion Detection with Computer Vision and Deep Learning. *Sensors*, *24*(11), 3484.
2. Gupta, S., Kumar, P., & Tekchandani, R. K. (2023). Facial emotion recognition based real- time learner engagement detection system in online learning context using deep learning models. *Multimedia Tools and Applications*, *82*(8), 11365-11394.
3. Joseph, C. W., Kathrine, G. J. W., Vimal, S., Sumathi, S., Pelusi, D., Valencia, X. P. B., & Verdú, E. (2024). Improved optimizer with deep learning model for emotion detection and classification. *Mathematical biosciences and engineering: MBE*, *21*(7), 6631-6657.
4. Jagadeesh, M., & Baranidharan, B. (2022). Facial expression recognition of online learners from real-time videos using a novel deep learning model. *Multimedia Systems*, *28*(6), 2285-2305.
5. Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V., & Le, Q. V. (2019). Autoaugment: Learning augmentation strategies from data. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 113-123).
6. Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence*, *39*(6), 1137-1149.
7. Michalski, B., & Plechawska-Wójcik, M. (2022). Comparison of LeNet-5, AlexNet and GoogLeNet models in handwriting recognition. *Journal of Computer Sciences Institute*, *23*, 145-151.
8. What is deep learning? <https://www.mathworks.com/discovery/deep-learning.html>
9. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84-90.
10. Vaidya, K. S., Patil, P. M., & Alagirisamy, M. (2023). Hybrid CNN-SVM classifier for human emotion recognition using ROI extraction and feature fusion. *Wireless Personal Communications*, *132*(2), 1099-1135.
11. Wurzberger, F., & Schwenker, F. (2024). Learning in Deep Radial Basis Function Networks. *Entropy*, *26*(5), 368.
12. Ioffe, S. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv: 1502.03167*.
13. Liu, Y. Q., Du, X., Shen, H. L., & Chen, S. J. (2020). Estimating generalized gaussian blur kernels for out-of-focus image deblurring. *IEEE Transactions on circuits and systems for video technology*, *31*(3), 829-843.
14. Hu, C., Li, H., Ma, T., Zeng, C., & Ji, X. (2024). An improved image enhancement algorithm: Radial contrast-limited adaptive histogram equalization. *Multimedia Tools and Applications*, 1-13.

15. Understanding Digital Images for Image Processing and Computer Vision (Part 1), <https://medium.com/@md-jewel/understanding-digital-images-for-image-processing-and-computer-vision-part-1-cc42be78cca1>
16. Rathee, N., Pahal, S., & Sheoran, D. (2021). Evaluating the uncertainty of classification due to image resizing techniques for satellite image classification. *Mapan*, 36(2), 243-251.
17. Wu, J. (2017). Introduction to convolutional neural networks. *National Key Lab for Novel Software Technology. Nanjing University. China*, 5(23), 495.
18. Introduction To Pooling Layers In CNN, <https://pub.towardsai.net/introduction-to-pooling-layers-in-cnn-dafe61eabe34>
19. How ReLU and Dropout Layers Work in CNNs, <https://www.baeldung.com/cs/ml-relu-dropout-layers>
20. Zafar, A., Saba, N., Arshad, A., Alabrah, A., Riaz, S., Suleman, M., ... & Nadeem, M. (2024). Convolutional Neural Networks: A Comprehensive Evaluation and Benchmarking of Pooling Layer Variants. *Symmetry*, 16(11), 1516.
21. Vasyltsov, I., & Chang, W. (2021). Efficient softmax approximation for deep neural networks with attention mechanism. *arXiv preprint arXiv: 2111.10770*.
22. Understanding the Softmax Activation Function: A Comprehensive Guide, <https://www.singlestore.com/blog/a-guide-to-softmax-activation-function/>
23. Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv: 1412.6980*.
24. Li, S., & Deng, W. (2020). Deep facial expression recognition: A survey. *IEEE transactions on affective computing*, 13(3), 1195-1215.
25. Terven, J., Cordova-Esparza, D. M., Ramirez-Pedraza, A., Chavez-Urbiola, E. A., & Romero-Gonzalez, J. A. (2023). Loss functions and metrics in deep learning. *arXiv preprint arXiv: 2307.02694*.