

מוטיבציה לשימוש ברשתות קונבולוציה

ביחידות הלימוד הקודמות למדנו כיצד לאמן רשת נירונים לסיווג תמונות פריט הלבוש באוסף הנתונים Fashion-MNIST, אך לא השקענו מחשבה רבה בבחירת ארכיטקטורת הרשתות אותן אימנו. דבר זה לא היווה בעיה שכן שכבות הרשת תמיד היו בעלות קישוריות מלאה (FC - fully connected layers), בהן כל נירון מחובר לכל נירון בשכבה העוקבת. שכבות אלו בעלות כוח הבעה רב, ובחירה מתאימה של משקלי הרשת אפשרה להן ללמוד כללים המפרידים בין מחלקות פרטי הלבוש. יחד עם זאת, לכוח הבעה פוטנציאלי זה יש מחיר: השימוש בשכבות אלו יקר מבחינה חישובית, וכן יש בהן מספר רב של פרמטרים כך שדרוש סט אימון גדול במיוחד בכדי לאמן. כל עוד הבעיה בה אנו עוסקים פשוטה מספיק, כגון סיווג תמונות שחור-לבן בגודל 28×28 פיקסלים לעשר מחלקות, מחיר זה לא בא לידי ביטוי, אך מהר מאוד עם מורכבות הבעיה המחיר הופך בלתי ריאלי. חשבו לדוגמה על רשת אשר הקלט שלה הוא תמונות צבעוניות בגודל 1×1 מגה-פיקסל: מימד הקלט הוא 3 מיליון (מיליון פיקסלים לכל אחד משלושת ערוצי הצבע), ולכן **לכל נירון** בשכבה השנייה יהיו 3 מיליון משקלים. אפילו אם השכבה השנייה תהיה קטנה והרשת לא עמוקה במיוחד, העלות החישובית של מעבר קדימה יחיד תהיה גבוהה, על אחת כמה וכמה העלות של תהליך האופטימיזציה.

כוח ההבעה הרב של שכבות FC מאפשר גמישות, אך לעתים כזו שאין לנו צורך בה: חשבו למשל על כך שבכל הרשתות אשר אימנו עד כה, מיקום הפיקסלים בתמונה לא נלקח בחשבון. ייתכן שעם תהליך האימון הרשת למדה שקיימת קורלציה בין ערכיהם של פיקסלים מסויימים, אך היא יכלה ללמוד זאת גם אם הקלט היה "מעורבב" לפני הזנתו, וזאת מפני שבתחילת תהליך האימון שיטחנו את טנזור הקלט, וכל איבר של טנזור הקלט השפיע באופן זהה על החישוב העוקב. לצורך דוגמה, נגריל תמורה רנדומלית על שורות התמונות,

```
row_mix=torch.randperm(28)
print(row_mix)
```

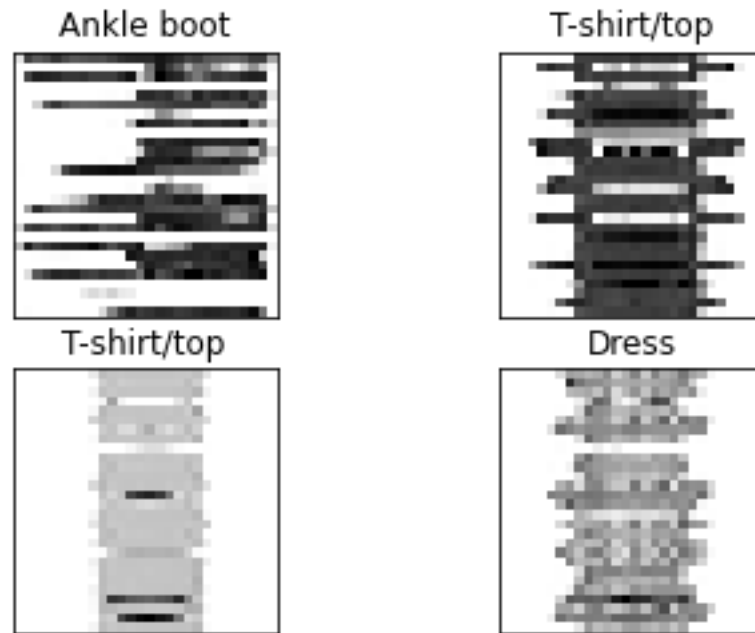
פלט:

```
tensor([21,  6, 18,  0, 11, 23,  4, 12, 27,  9,  7, 13, 24,  3,
        5, 16, 22,  8, 20, 26, 19, 14, 10, 17,  1, 25,  2, 15])
```

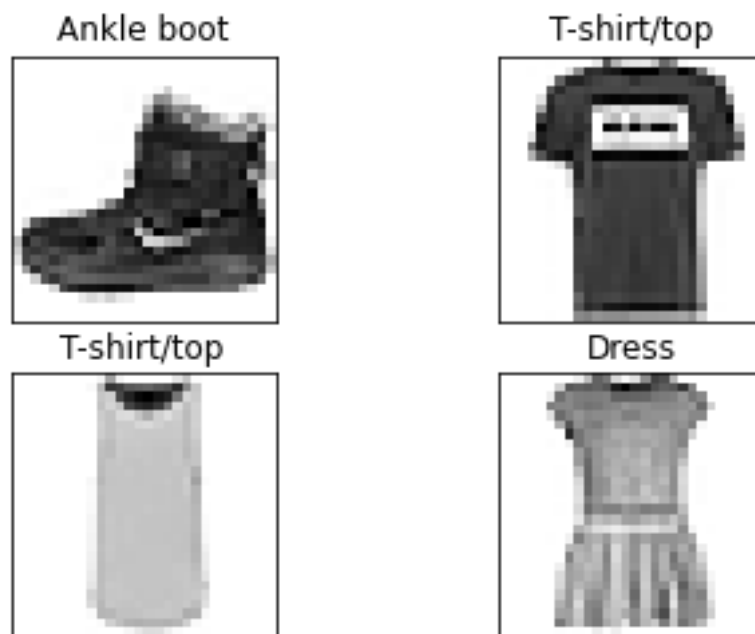
ונפעיל תמורה זו על השורות של כל התמונות בסט האימון בדרך הבאה.

```
imgs = imgs[:, :, row_mix, :]
```

זכרו שהמשתנה `imgs`, הנטען דרך ה-Dataloader, מכיל batch של תמונות ועל כן הוא טנזור 4 מימדי: המימד הראשון הוא מימד ה-batch, השלישי והרביעי הם מימדי השורה והעמודה של התמונה (ועל כן גודל כל אחד מהם הוא 28). המימד השני הוא מימד הערוץ (Channel), אך התמונות באוסף נתונים זה הן בשחור-לבן ולכן בעלות רק ערוץ יחיד. תמונות צבעוניות לרוב יכילו שלושה ערוצים עבור עוצמת האדום, הירוק והכחול בפיקסל הנתון. נפעיל טרנספורמציה זו על 4 פרטי לבוש מקריים ונאייר זאת להלן.



כאשר פרטי הלבוש המקוריים הם:



בעוד שלעין אנושית כלל לא ברור מהאיור הראשון מהם פרטי הלבוש הרלוונטיים, רשתות הנוירונים בהן עסקנו עד כה יגיעו לביצועים זהים עבור כל אחת מצורות הקלט, כל עוד נפעיל את אותה הטרנספורמציה על כל התמונות באוסף נתוני האימון והבדיקה.

ביחידה הנוכחית נוותר על עוצמה זו, אך בתמורה נקבל שכבות יעילות בהרבה המותאמות באופן ספציפי למשימות העוסקות בעיבוד תמונה וזיהוי פרטים. יעילות זו תאפשר לנו לאמן רשתות עמוקות יותר, אשר לבסוף, בעבור אותו מחיר חישובי, יבצעו עבודה טובה בהרבה מרשתות בעלות קישוריות מלאה. הרשתות החדשות יורכבו משכבות קונבולוציה (Convolution), הלוקחות השראה ממערכת הראיה האנושית ומשיטות עיבוד תמונה דיגיטלי קלאסיות. שכבות קונבולוציה מסכמות את השכבה הקודמת על ידי חילוץ מאפיינים פשוטים (כגון קיום פינה במיקום מסוים בתמונה). על ידי שרשרון אחת לאחר השניה באופן היררכי, מחולצים לעומק הרשת מאפיינים מורכבים יותר של התמונה המקורית (כגון קיומם של שרוול חולצה או שרוכי נעליים בתמונה). לבסוף על בסיס המאפיינים שהרשת למדה לחלץ מהתמונה יתבצע הסיווג למחלקות השונות.

בהמשך יחידה זו תלמדו את הפרטים המדויקים של פעולת רשת כזו, הרכיבים השונים מהן היא מורכבת, וכמובן מימושן בקוד ושילובן באלגוריתם למידה מבוסס גרדיאנט.

שאלות לתרגול

1.

- א. אמנו רשת עמוקה בעלת קישוריות מלאה לחיזוי המחלקות של אוסף הנתונים Fashion-MNIST לאחר שהגרלתם תמורה אקראית **על כל הפיקסלים** בתמונה והפעלתם אותה על כל תמונה לפני הזנתה.
 - ב. בדקו את ביצועי הרשת על סט הבדיקה לאחר שעל כל תמונה בו הופעלה אותה תמורה.
 - ג. בדקו את ביצועי הרשת על סט הבדיקה כאשר לא הופעלה על התמונות בו כל תמורה.
2. אמנו שוב את הרשת כפי שעשיתם בשאלה 1, אך כעת הגרילו מחדש תמורה אקראית עבור כל minibatch. נסו להסביר את התוצאה המתקבלת.