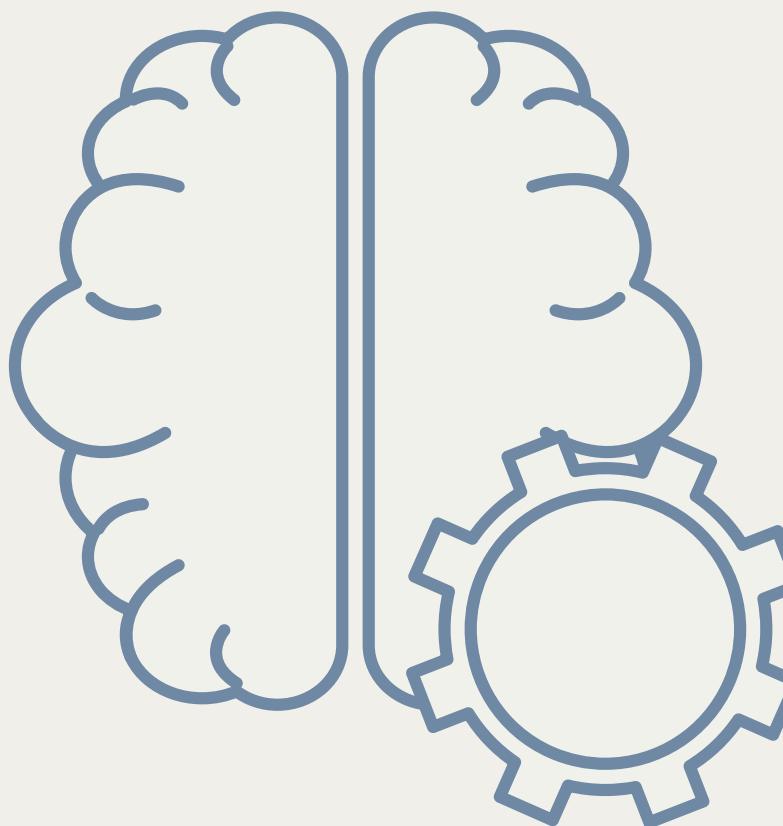


# Improving MLP-Mixer Performance on Small Image Datasets Using Locality-Aware Modifications

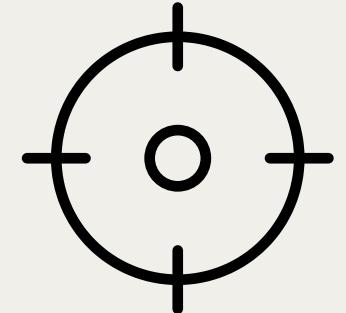
שיפור ביצועים על מאגרי תМОנות קטנים באמצעות  
הוספת הטיה מרחבית

---

מגישים: גורא קנאזע, עמר מנאו, אדם עיסא



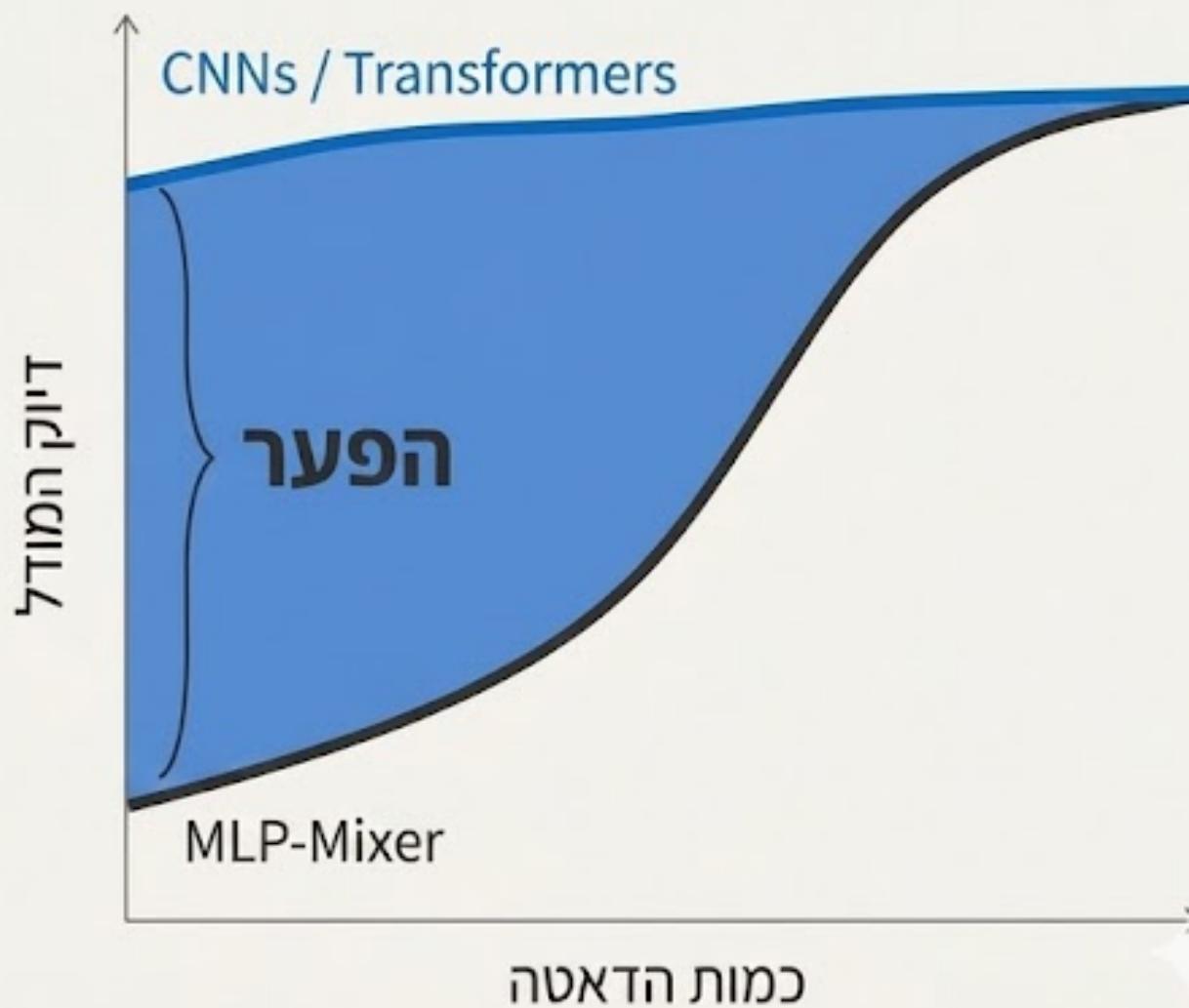
## **מטרת הפרויקט: שיפור הדיווק של ארכיטקטורת MLP-Mixer על מאגרי מידע קטנים (CIFAR-10/100)**



**הגישה המרכזית:** החדרת הטיה אינדוקטיבית (inductive bias) מרחבית מינימלית באמצעות שכבה קונבולוציה יעה (convolution point wise & depth) שלא מפירה את המוטיבציה של מבנה הרשת וambil להוסיף כמות גדולה של פרמטרים נוספים.

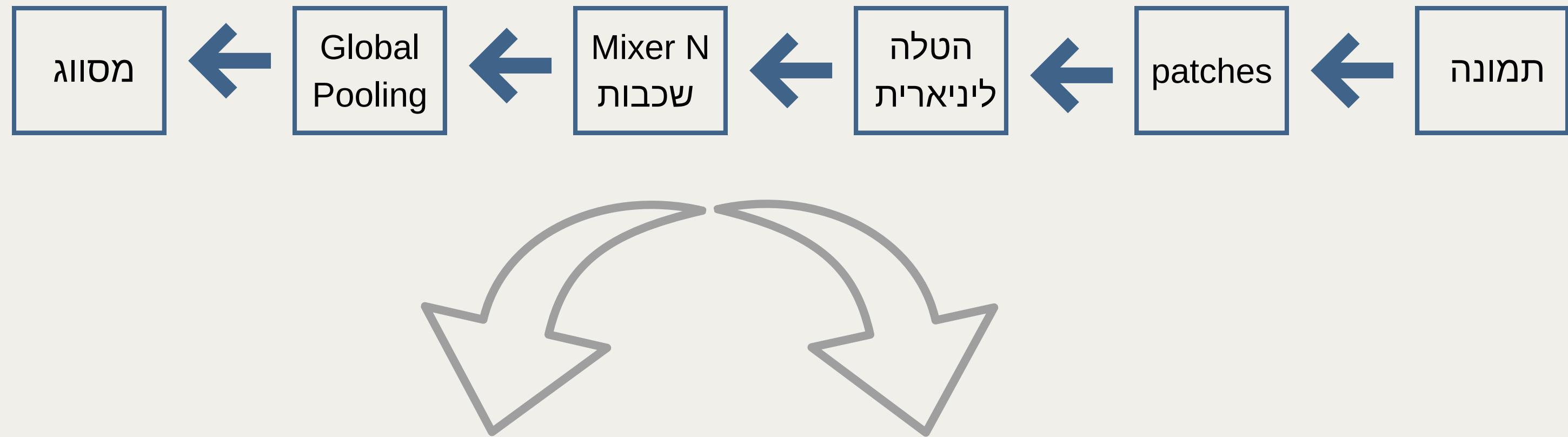


# המוטיבציה: הפער בין ביצועים על דатаה גדול לקטן



- **הבטחה של MLP-Mixer:** ארכיטקטורה חדשה שמתחרה ב-CNNs ובל-Transformer ללא שימוש בקונבולוציות או attention.
- **המציאות:** המודל משיג ביצועים מעולים רק כאשר הוא מאומן על מאגרי מידע עצומים (כמו W300-JT).
- **הבעיה:** על מאגרי מידע קטנים (כמו CIFAR), ביצועיו נמוכים משמעותית ממודלים קלאסיים.
- **הזהדנות:** קיים פער ברור המזמין שיפורים מוקדים כדי להפוך את המודל לרבונטי גם בסביבות דלות-מידע.

# כיצד MLP-Mixer עובד?



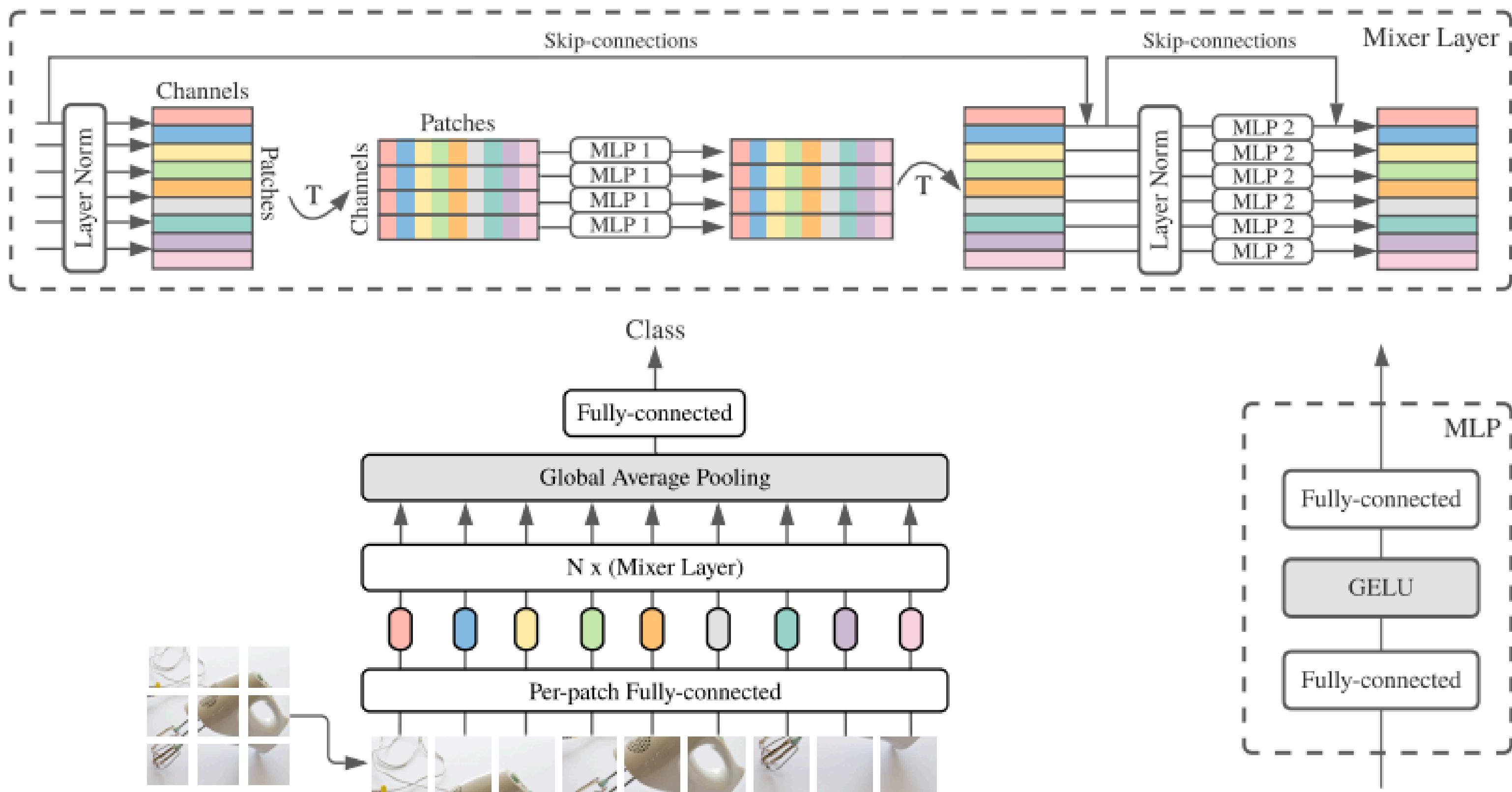
## Token-Mixing MLPs .2

- רשת MLP שמערבתת מידע בין הטלאים בתמונה.
- מאפשרת תקשורת מרחבית בין אזורים שונים.
- נקודת חשובה: הערבות הוא גלובלי, בלי "חוק שכנות" מובנה.

## Channel-Mixing MLPs .1

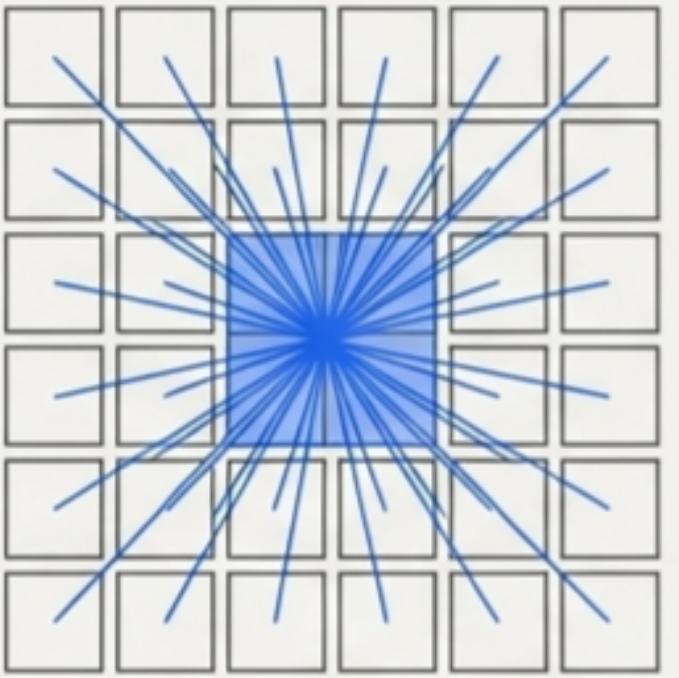
- רשת MLP שפועלת על כל טלאי בנפרד.
- מעורבתת תכונות/ערוצים בתוך אותו טלאי.
- אינטואיציה: כמו "עיבוד תכונות" מקומי בתוך הטלאי.

# ארQUITקטורה מלאה של הרשת



# מדוע הידר לוקאליות פוגע ביצועים על מידע מועט?

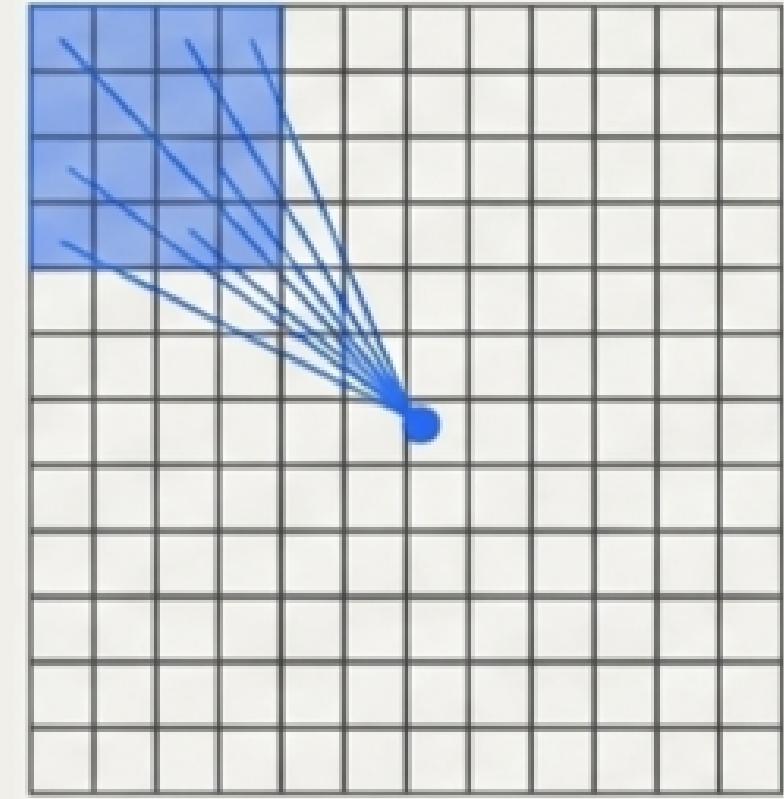
(Learning from Scratch - למידה מאפס (MLP-Mixer



- ה-Mixer "ראה" את כל התמונה בביטחון אחת, אך אינו יודע אילו טלאים שכנים.
- הוא צריך ללמד את מושג ה'שכנות' והמבנה המרחבי מתוך כמות עצומה של דוגמאות.

**בלי מספיק נתונים, ה-Mixer מתנסה ללמידה את הקשרים המרחביים ש-NNC מקבל ב"מתנה"**

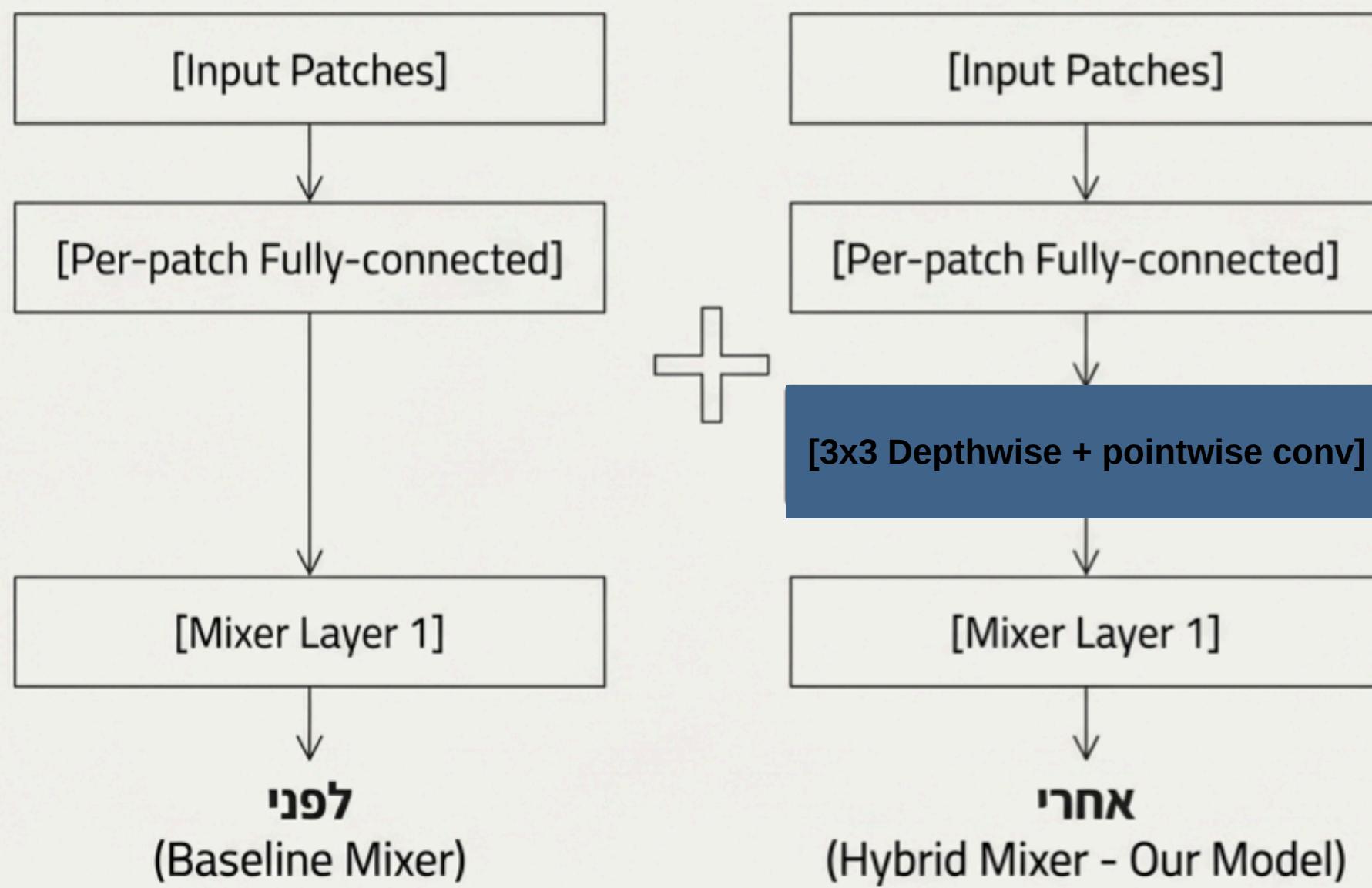
(Built-in Bias - הטיה מובנית (CNN



- CNN מתוכננים מראש עם ההנחה שפיקסלים קרובים קשורים זה לזה.
- הפילטר לומד תבניות מקומיות (קצוות, טקסטורות) באופן טבעי ויעיל.
- הנחה זו (הטיה אינדוקטיבית) חוסכת למודל את הצורך "ללמידה" את חשיבות המבנה המרחבי מאפס.

# הרעין שלנו : הזרקת לوكאליות בשינוי מינימלי

Before vs. After



השינוי:

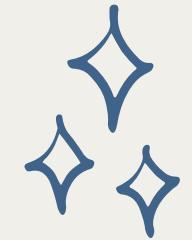
אנו מוסיפים שכבה קונבולוציה Depthwise בגודל  $3 \times 3$   
ואז מפעילים קונבולוציה Pointwise בגודל  $1 \times 1$ , מיד  
אחרי הטמעת הטלאים (patch embedding).

ההיגיון:

הפעלת קונבולוציה عمוקית (Depthwise) ולאחריה  
קונבולוציות נקודתיות (Pointwise) מאפשרת תחילת  
 לכל טלי או לתקשר עם שכניו המרחביים הקרובים,  
 ובשלב הבא לשלב מידע בין ערכאים, וכך ליצור ייצוג  
ראשוני בעל מודעות מרחבית וסמנטית לפני הכניסה  
 לבлокי ה-Mixer האלובליים.

# מטרות הפרויקט והשערת המחקר

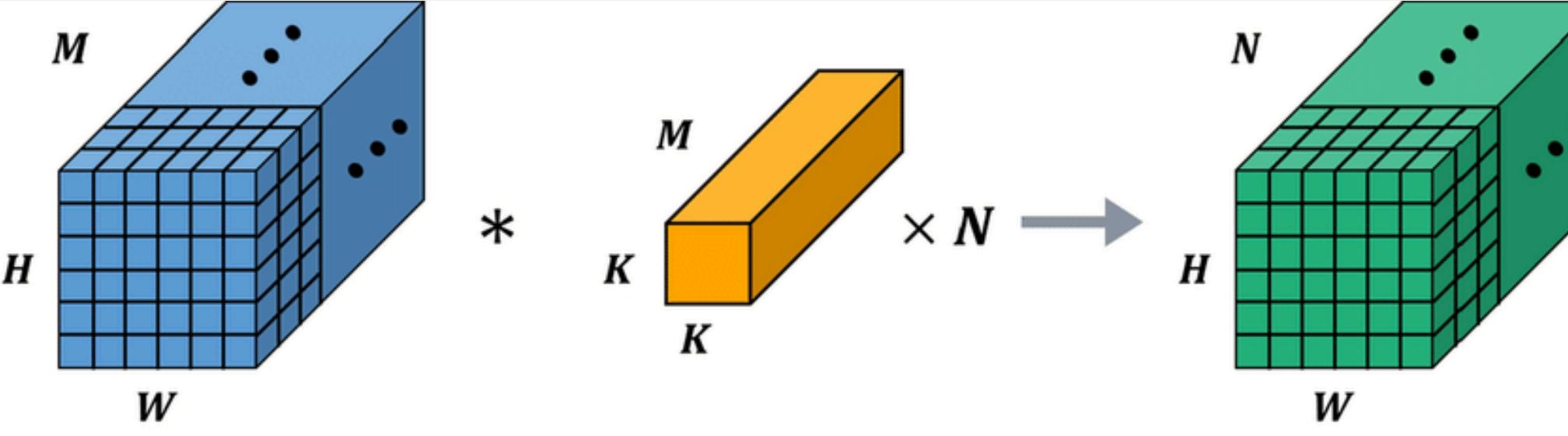
מטרה מרכזית לשפר את רמת הדיוק (accuracy) ויכולת ההכללה (generalization) של ארכיטקטורת MLP-Mixer על מאגרי מידע קטנים (כדוגמת 10/100 CIFAR-10).



**השערת המחקר :** תוספת של שכבה קונבולוציה קלת-משקל (lightweight) תספק למודל-h-Mixer את הנטיה המרחבית החסירה לו, ותובייל להפחית התאמת-יתר (overfitting) ולעלייה ביצועי הכללה במצב של מיעוט נתונים.



## Standard Convolution



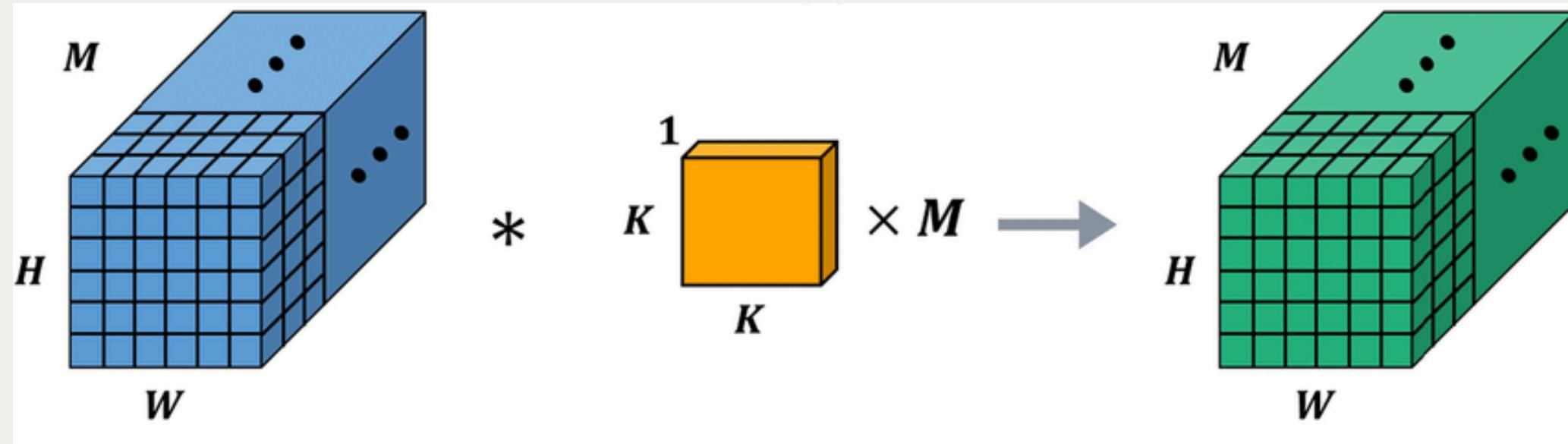
## Standard Vs Depthwise Vs Pointwise Convolution

# of params: Kernel\_size \* Kernel\_size \*  
in-channels \* out-channels

$$3 \times 3 \times 128 \times 128 = 147,456$$

## Depthwise Convolution

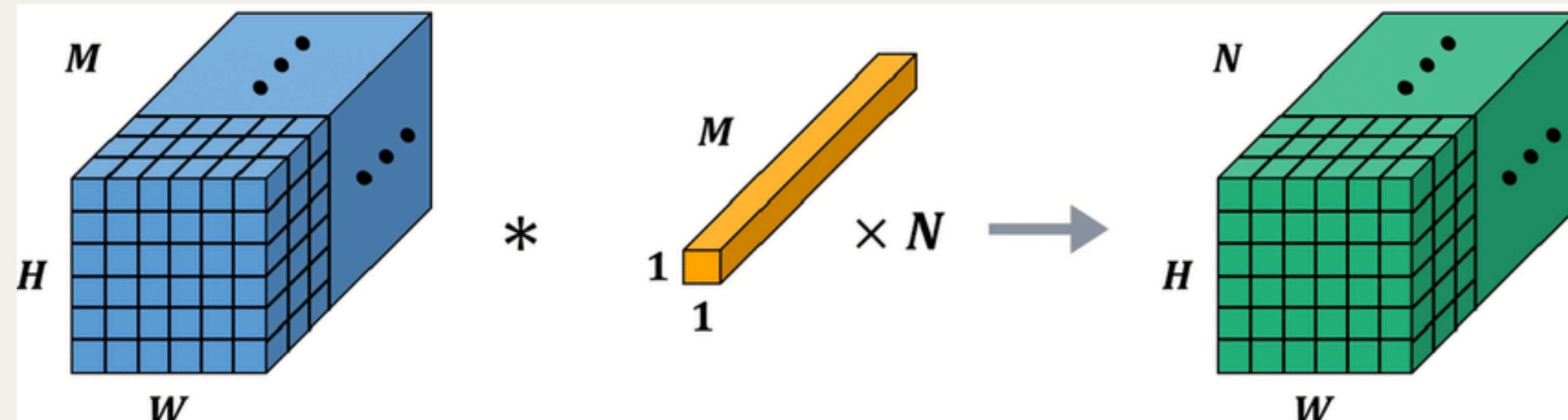
# of params:  
Kernel\_size \*  
Kernel\_size \*  
channels



$$3 \times 3 \times 128 = 1,152$$

# of params:  
channels \*  
channels

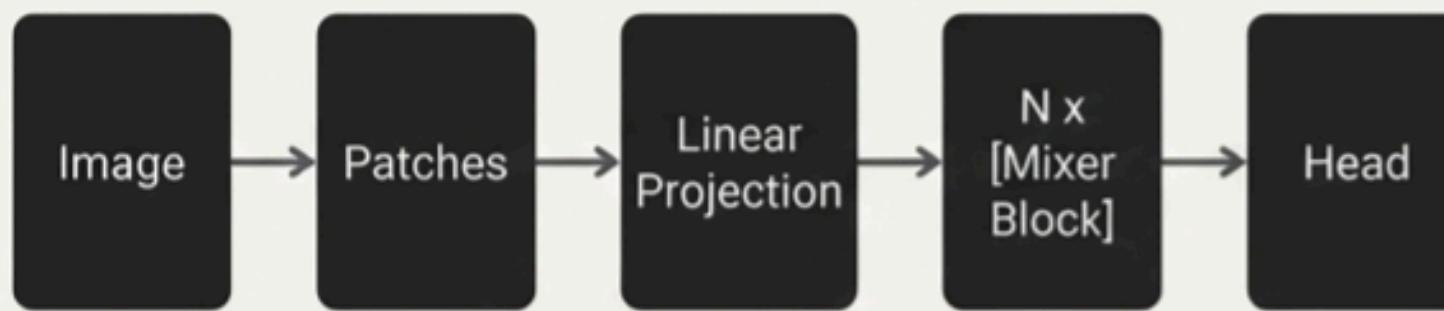
## Pointwise Convolution



$$128 \times 128 = 16,384$$

# השוואת ארכיטקטורות: Baseline VS Hybrid

## MLP-Mixer (Baseline)



## Hybrid Mixer (Our Model)



**שמירה על הצורה (Shape):**  
הكونבולוציה אינה משנה את  
מספר הטעוקנים או המימד  
שליהם, ומשתלבת באופן  
סקופי.

**שינוי יחיד וממוקד:** התוספת  
מרתחשת לפני שכבות ה-  
Mixer, ומכינה עבורן ייצאים  
'מודעי-רחבי'.

# מארן הנתונים

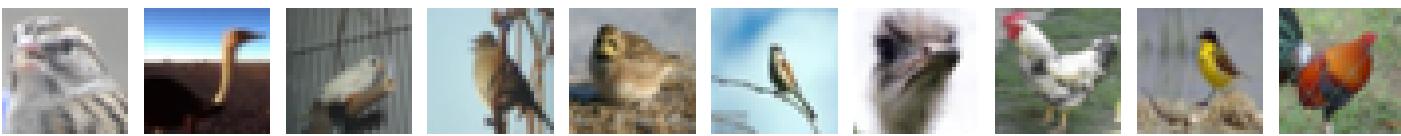
**airplane**



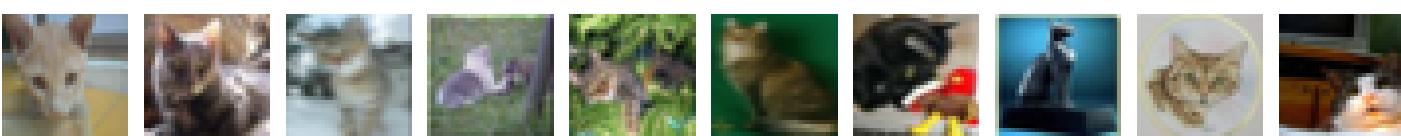
**automobile**



**bird**



**cat**



**deer**



**dog**



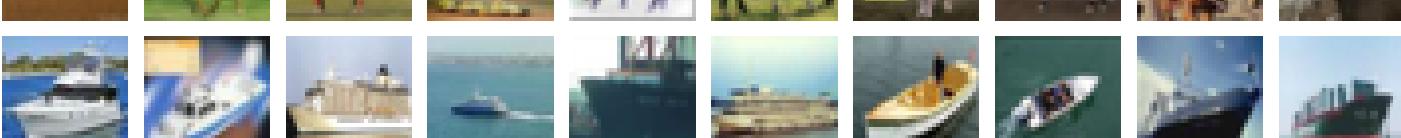
**frog**



**horse**



**ship**



**truck**



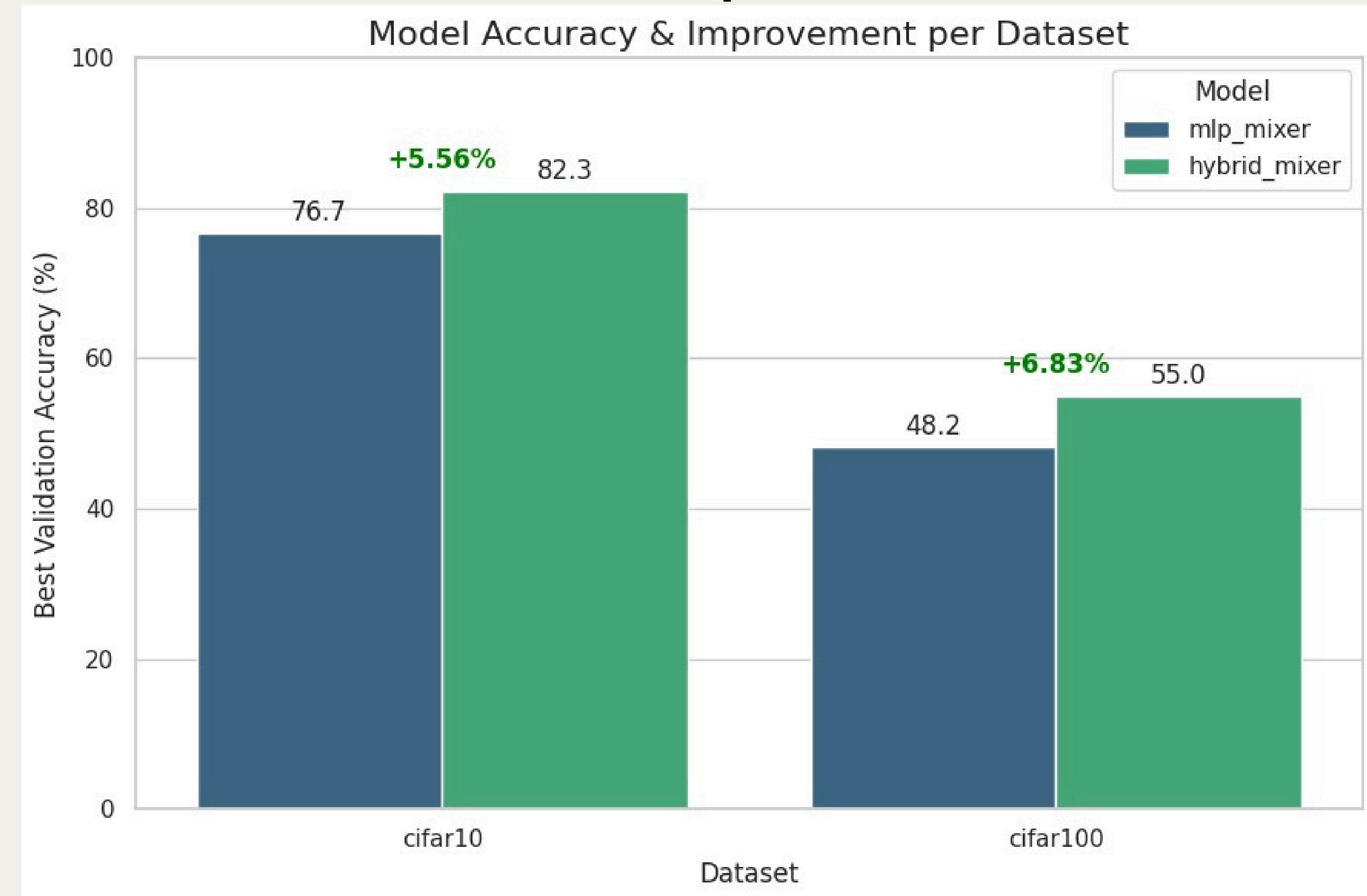
## ראשי: CIFAR-10

- מארן סטנדרטי להשוואת מודלים ב-10 קטגוריות.
- גודלו המוגבל אידיאלי לבחינת ההשערה שלנו.

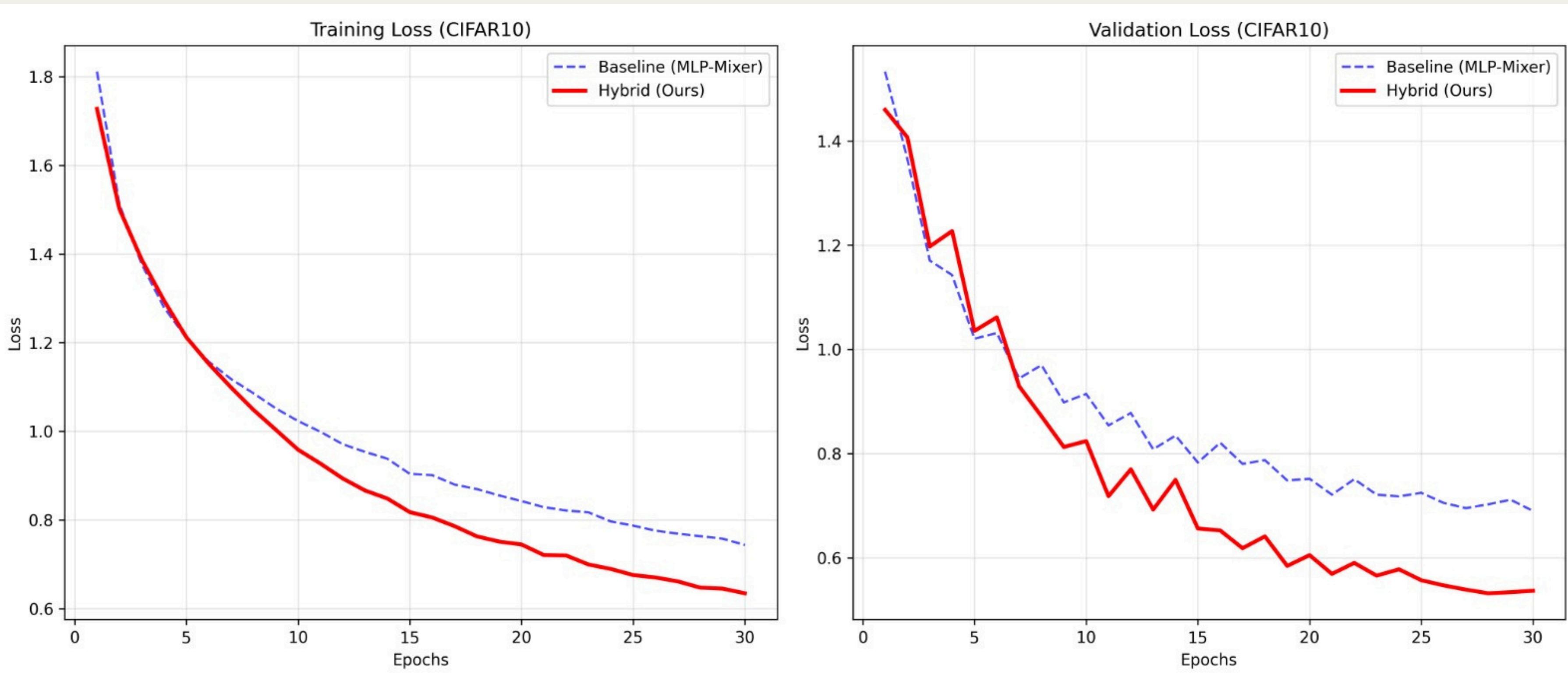
## משני: CIFAR-100

לבחינת יכולת הכלכלה על בעיה מורכבת יותר.

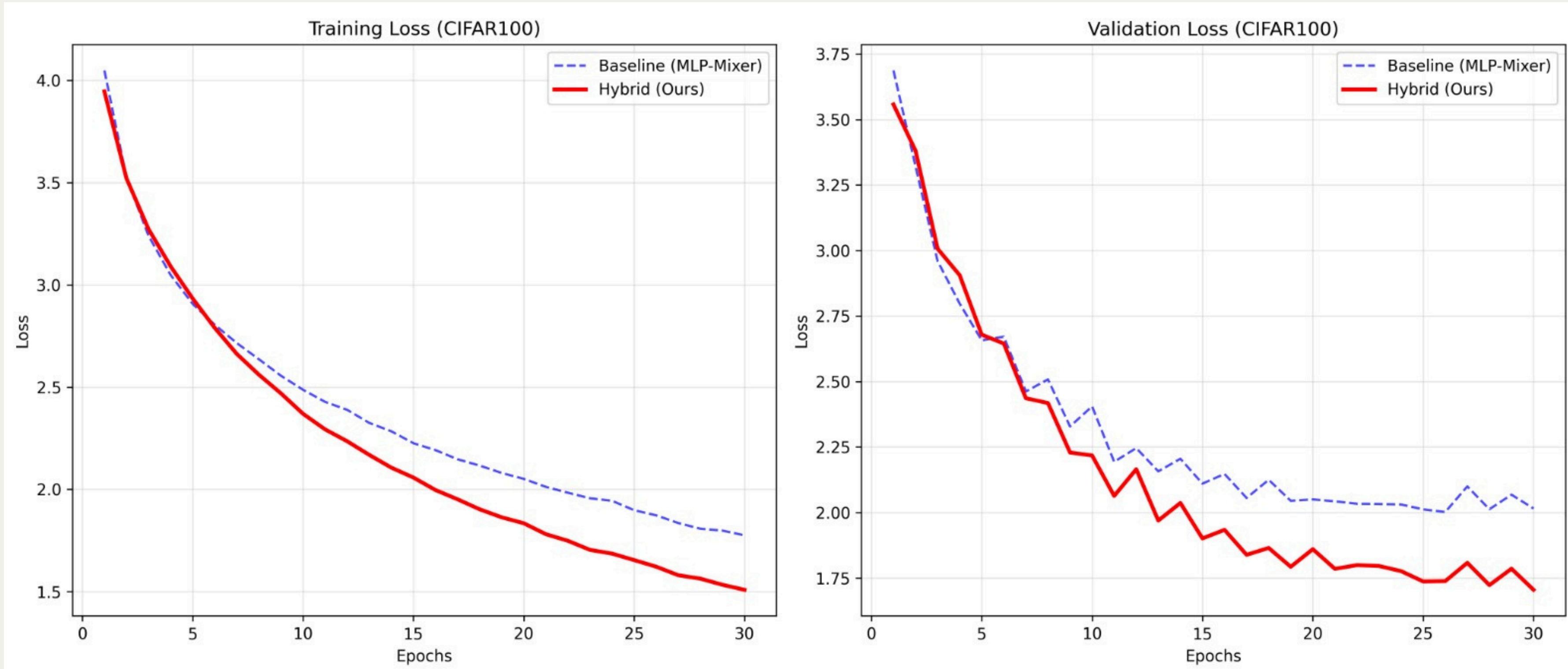
# תוצאות: שיפור משמעותי בדיק הSION



# השוואה ביצועים של המודלים (cifar 10)



# השוואת ביצועים של המודלים (cifar 100)



# מגבלות

נבדקה רק תוספת של  $W \times 3$  Conv. יתכנו מנגנוני הטיה אחרים (למשל  $5 \times 5$ , או Pooling).

הניסויים נערךו רק על מאגרי מידע קטנים (CIFAR). לא ידועה ההשפעה על מאגרים גדולים יותר כמו ImageNet.

