HIL-UMI entrena robots sin depender de la máquina real

20 de octubre de 2026, 雷峰网 (Leiphone). Selección: Miguel Acevedo

ia-fisica, aprendizaje-por-imitacion, investigacion, china, primebot



Últimamente, la industria de la IA física discute con fuerza si se puede hacer el postentrenamiento sin el cuerpo del robot. Dentro y fuera de China, todavía ningún equipo había mostrado un esquema viable.

Dicho en simple: ni un modelo grande ni un robot de IA física se vuelven realmente útiles solo con el preentrenamiento. El aterrizaje depende del postentrenamiento, el ajuste que viene después. Hao Dong, científico jefe de PrimeBot (启元机器人), dijo a AI Tech Review: “El preentrenamiento sube el piso. El techo lo define el postentrenamiento.”

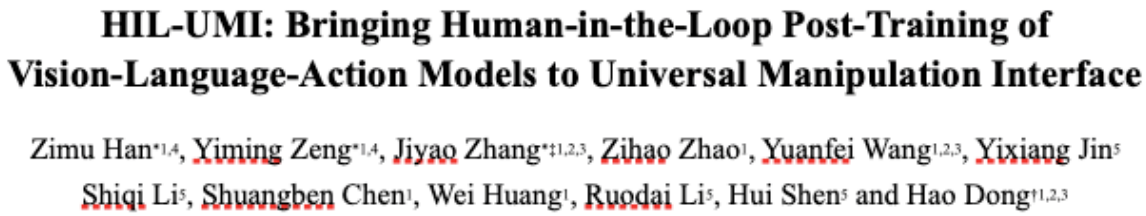
Las empresas de IA física quieren entrar a la fábrica, limpiar una casa y otros usos reales. Lo que se compara es la tasa de éxito, o sea la fuerza del postentrenamiento.

La dificultad es que ese ajuste tiene que tratar con el mundo físico. Hasta ahora se hacía casi solo con datos del robot real.

Entrenar en la máquina real sale carísimo. Para escalar harían falta muchísimos cuerpos de robot, y el desgaste de hardware, el espacio y la mano de obra cuestan una fortuna. Por eso el preentrenamiento ya empezó a recoger datos con dispositivos sin cuerpo. El volumen subió, pero el eslabón más importante del postentrenamiento tradicional, la corrección de errores, quedó sin resolver.

Para ese problema, la Universidad de Pekín y el equipo de Hao Dong en PrimeBot (启元机器人) presentaron HIL-UMI.

El trabajo se concentra en el postentrenamiento de IA física sin el cuerpo del robot, y en particular en esa corrección.



HIL-UMI: Bringing Human-in-the-Loop Post-Training of Vision-Language-Action Models to Universal Manipulation Interface

Zimu Han[*1,4], Yiming Zeng[*1,4], Jiyao Zhang[*‡1,2,3], Zihao Zhao[1], Yuanfei Wang[1,2,3], Yixiang Jin[5]
Shiqi Li[5], Shuangben Chen[1], Wei Huang[1], Ruodai Li[5], Hui Shen[5] and Hao Dong[†1,2,3]

Artículo: https://arxiv.org/pdf/2609.20659

Página del proyecto: https://hil-umi.github.io

El costo del postentrenamiento de IA física

Por qué supera por mucho al RLHF de los modelos grandes

Frente a los modelos grandes, el postentrenamiento de IA física es tan caro que muchos modelos de robots ni siquiera pueden intentar escalar.

Que un modelo de lenguaje sea tan útil no se explica solo por el volumen del preentrenamiento. También pesan el RLHF (aprendizaje por refuerzo con retroalimentación humana, en simple: el etiquetado de datos) y métodos de postentrenamiento como DPO.

En un robot, ese ajuste cuesta bastante más de cien veces el RLHF de un modelo grande. Para el RLHF de un modelo de lenguaje alcanza con sentar a más personas frente a una computadora a elegir opciones. Un robot de IA física tiene que vérselas de verdad con el mundo físico.

El método tradicional más usado, DAgger, pone a un humano o a un algoritmo como experto que evalúa el postentrenamiento. Funciona, pero es caro y lento.

La forma más común es el humano en el bucle sobre la máquina real, o HG-DAgger (Human-Gated DAgger). El brazo trabaja y un operador mira, muy atento, sentado al lado. Cuando el robot está por agarrar al vacío o por chocar, toma el control y graba la trayectoria correcta. El humano es quien cierra todo el circuito de entrenamiento.

Lo más engorroso es que, después de cada corrección, alguien tiene que dejar la escena como estaba. El robot ejecuta, la persona interviene, el entorno se reinicia. Una tarde entera deja pocos datos útiles.

La mayor parte de ese tiempo se va en los movimientos torpes del robot y en rearmar la escena a mano.

Ese modo, lento y caro, convirtió la escala de los datos de postentrenamiento en un problema crónico y limitó con fuerza el crecimiento.

De ahí sale la pregunta central: ¿se puede dejar de lado la máquina real y, como quien hace las tareas de la casa, recoger de paso los datos de corrección que el modelo más necesita?

La lógica de fondo de HIL-UMI

Un cerebro de modelo para la captura con UMI

El método de HIL-UMI se resume en una frase: se puede dar clase de recuperación al robot sin usar un robot.

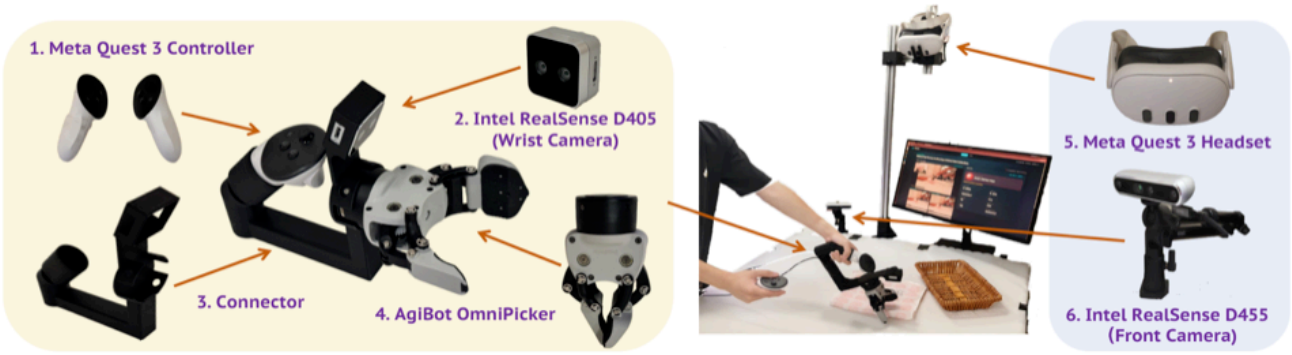
HIL-UMI es UMI con humano en el bucle (Human-In-the-Loop). A diferencia del UMI habitual, la pinza de mano con la que se recogen datos, aquí la captura no es ciega. Entra un modelo.

HIL-UMI conecta el modelo directamente a esa captura de mano:

En el mundo real solo se mueve la persona. El operador trabaja con la pinza UMI, por ejemplo agarra un juguete o dobla una toalla, y la cámara graba la escena al mismo tiempo.

El modelo imagina en segundo plano. Recibe las mismas imágenes y la misma instrucción, y predice rápido cómo serían los siguientes 20 pasos si actuara él. No controla ningún brazo.

Si hay divergencia, salta un aviso. El sistema compara en tiempo real la acción humana y la acción predicha. Si se parecen, el modelo ya lo domina y no hace falta volver a grabar. Si la trayectoria que el modelo imagina se desvía mucho de lo que hace la persona, señal de que en esa escena todavía es un novato, el sistema avisa de inmediato: este tramo el modelo no lo sabe, desde aquí enséñale un poco más.



Pinza de mano y equipo de captura.

Todo el proceso ocurre sin un brazo robótico en la sala y sin que el robot tenga que equivocarse de verdad. Mientras demuestra, la persona captura justo las escenas en las que el modelo, ya desplegado, tiene más probabilidad de fallar.

No es solo una herramienta de captura

Es un ciclo de datos del postentrenamiento

Si HIL-UMI se lee apenas como una grabadora de mano, se lo subestima. Lo fuerte es un sistema cerrado que hace que el modelo mismo saque su cuaderno de errores.

Ese ciclo tiene dos pilares técnicos.

Primer paso: detección de divergencia de acciones, para encontrar los puntos ciegos. Ante una misma situación, un robot suele tener varias posturas razonables. HIL-UMI no mira una sola predicción. Muestrea en paralelo varios bloques de acción y calcula una distribución de energía (Energy Score). Si el modelo está seguro, la distribución se concentra. Si las acciones salen dispersas y además se alejan mucho de las de la persona, el sistema juzga que entró en un estado fuera de distribución (OOD) y arranca la grabación extra.

Segundo paso: un estimador de ventaja, para separar la acción útil del movimiento sin sentido. Saber moverse no alcanza. Hay que saber si el movimiento hace avanzar la tarea. En apilar bloques, dudar en el aire con la pieza agarrada y alinearla con precisión sobre la base son dos cosas distintas. HIL-UMI incorpora un estimador que, a partir de la imagen anterior y la posterior, mide cuánto avanzó la tarea. El entrenamiento queda condicionado por esa ventaja: el modelo no solo aprende la corrección, sino que prioriza las acciones que de verdad empujan el progreso.

El cierre real está en la actualización dinámica. Después de cada ronda de datos dirigidos, el modelo itera. En la captura siguiente ya es más capaz, y los puntos que antes lo trababan dejan de disparar la alarma. La divergencia con la persona se corre hacia escenas de borde más profundas y más difíciles.

El modelo va mostrando puntos ciegos nuevos, la persona demuestra justo ahí, y el modelo vuelve a cambiar. La línea se itera sola, como una bola de nieve, y deja atrás la ineficiencia de grabar a ciegas enormes cantidades de ejemplos correctos.

Comparación de rutas de postentrenamiento sin robot

Para entender HIL-UMI con más frialdad, conviene ponerlo al lado de las rutas principales de la industria.

Teleoperación en robot real y DAgger en robot real: los datos son los más fieles, pero el costo por máquina es alto, la depreciación es grande y la barra de operación también. Llegar a

decenas de miles de unidades es casi imposible.

Simulación (Sim2Real) y aprendizaje por refuerzo: el cómputo es barato y se pueden generar muchos datos, pero con objetos flexibles, como la ropa, con la fricción y con iluminación compleja, la brecha entre simulación y realidad sigue siendo enorme. Además, no puede tomarse como postentrenamiento en la máquina real.

Preentrenamiento con videos masivos: la escala no tiene techo, pero faltan la información de fuerzas físicas y una dimensión precisa de la acción. Sirve sobre todo para dar al modelo una percepción de sentido común. No se traduce directo a un control milimétrico y se usa, en general, en el preentrenamiento.

UMI tradicional, captura positiva sin cuerpo: es barato y cualquiera puede recoger datos, pero solo junta ejemplos correctos. No resuelve cómo se recupera el modelo de los errores de cola larga, y también queda del lado del preentrenamiento.

HIL-UMI cae en un punto justo. Toma del UMI el bajo costo y la facilidad para escalar, toma de DAgger la corrección dirigida, y se saca de encima los defectos de los dos.

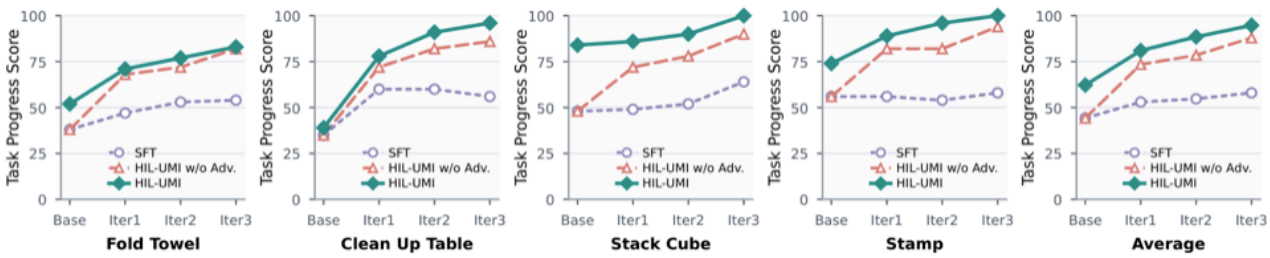
La eficiencia sube 5,63 veces

El postentrenamiento ya no tiene que girar alrededor del brazo

Si las cuentas de ingeniería cierran, se ve en dos comparaciones.

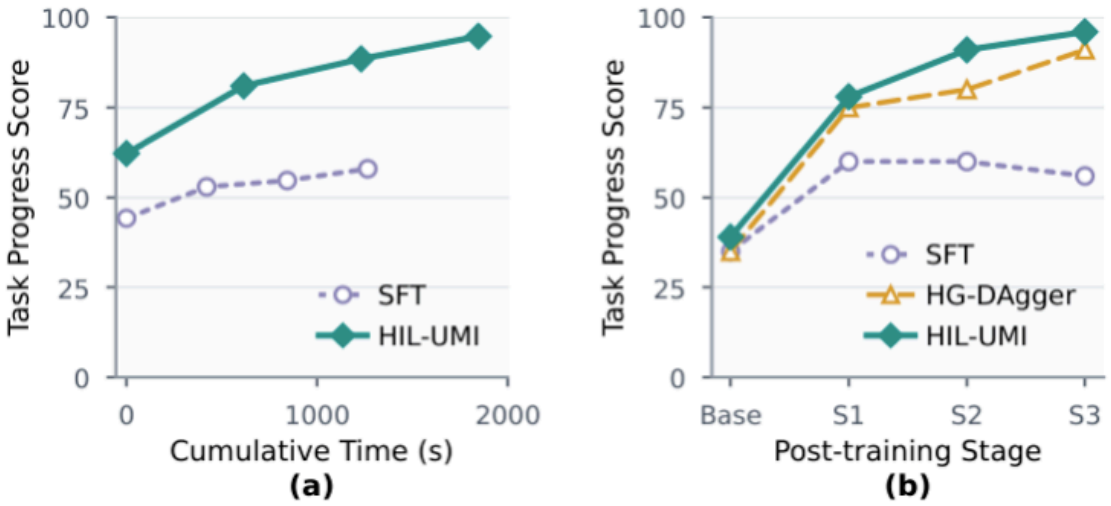
Clase dirigida contra aprender más a ciegas. En cuatro tareas con un brazo real (doblar una toalla, ordenar un escritorio, apilar bloques y estampar un sello), y con el mismo presupuesto de cuadros de datos, seguir sumando demostraciones comunes con un ajuste fino supervisado (SFT) apenas movió la puntuación media de avance de la tarea (TPS) de 44,3 a 58,0. Con la corrección dirigida de HIL-UMI, el promedio llegó a 94,8. En apilar bloques y estampar un sello, las dos tareas que más exigen precisión, la puntuación de éxito llegó a 100.

La eficiencia de captura se multiplicó por 5,63.



En el gráfico, el eje horizontal va del modelo inicial (Base) a tres rondas de actualización, de Iter1 a Iter3. El eje vertical es la puntuación de avance de la tarea. El violeta es SFT, el naranja es HIL-UMI sin el mecanismo de ventaja y el verde es HIL-UMI completo. A la derecha, el promedio de las cuatro tareas.

En la tarea larga y compleja de ordenar un escritorio, HG-DAgger en robot real tarda 412,99 milisegundos en capturar un cuadro útil. HIL-UMI tarda 73,40 milisegundos. La eficiencia por unidad de tiempo sube 5,63 veces, porque recoger con la mano es mucho más cómodo que teleoperar un robot.



A la izquierda, la puntuación media de las cuatro tareas según el tiempo acumulado de captura. A la derecha, el desempeño de cada método en ordenar el escritorio, según la etapa de entrenamiento.

El tiempo que se ahorra, más de cinco veces, es el que antes se iba en el brazo moviéndose despacio, en el operador esperando y en rearmar la escena a mano.

Para el medio, la imaginación comercial importa todavía más. La teleoperación del robot real exige una mano muy firme: un temblor chico ya deja datos malos. Con la pinza de mano, la persona ya sabe hacerlo y la calidad de la trayectoria es muy alta. Por primera vez, el postentrenamiento puede ser distribuido.

En el futuro, personal de medio tiempo en viviendas comunes de todo el país podría hacer gestos cotidianos con un equipo simple de unos cientos de yuanes, conectado al mismo modelo grande en la nube. Cuando el modelo detecta una divergencia de acciones en alguien, toma en silencio ese tramo corto de datos diferenciales de alta calidad y lo manda de vuelta.

El postentrenamiento de IA física, por fin, ya no exige que todo el mundo esté en el laboratorio terminando el trabajo con un brazo robótico.

El postentrenamiento que viene

Se va a jugar en velocidad y eficiencia

En la entrevista, Hao Dong, científico jefe de PrimeBot (启元机器人), dio un juicio que va contra el consenso:

> “El preentrenamiento y el postentrenamiento no son una simple relación de proporciones. Son un relevo por etapas.”

Para él, el caso se parece al de DeepMind con AlphaGo. Al principio usó partidas grabadas para aprendizaje por imitación, el preentrenamiento. Lo que de verdad lo consagró fue la etapa posterior: jugar contra sí mismo y hacer aprendizaje por refuerzo a gran escala. Con los modelos grandes de robots pasa lo mismo. El preentrenamiento arma la visión del mundo con la que se arranca, y esa ruta está relativamente clara. Quién llegue de verdad a la generalización y a una tasa de éxito extrema se va a decidir en un postentrenamiento que se pueda escalar.

Si los últimos dos años de IA física se jugaron a ver quién encendía primero un robot con un modelo grande, la segunda mitad se juega a ver quién arma, con el menor costo marginal y a la mayor velocidad, la red de corrección del robot.

En el giro que lleva a los modelos de IA física de una demostración que funciona a entrar en decenas de millones de casas, ¿qué cálculo está haciendo la industria? ¿Y cómo van a evolucionar la próxima generación de modelos base y los sistemas de postentrenamiento?

En breve, el medio publicará una entrevista exclusiva con Hao Dong, profesor asociado de la Universidad de Pekín y científico jefe de PrimeBot (启元机器人), sobre cómo se cruzan y se complementan la academia y la industria de la IA física.



Fuente: 雷峰网 (Leiphone)

roboticaasialatam.com/2026-10-20-hil-umi-ecea94/