



Autosaved just now

Edits visible only to you

Publish to project

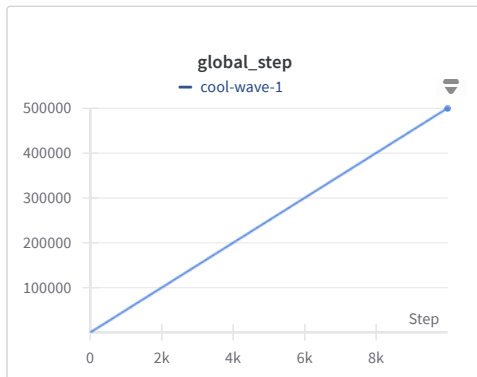


Wandb

Add a description...

Marouane BENFATTOUM

Section 1

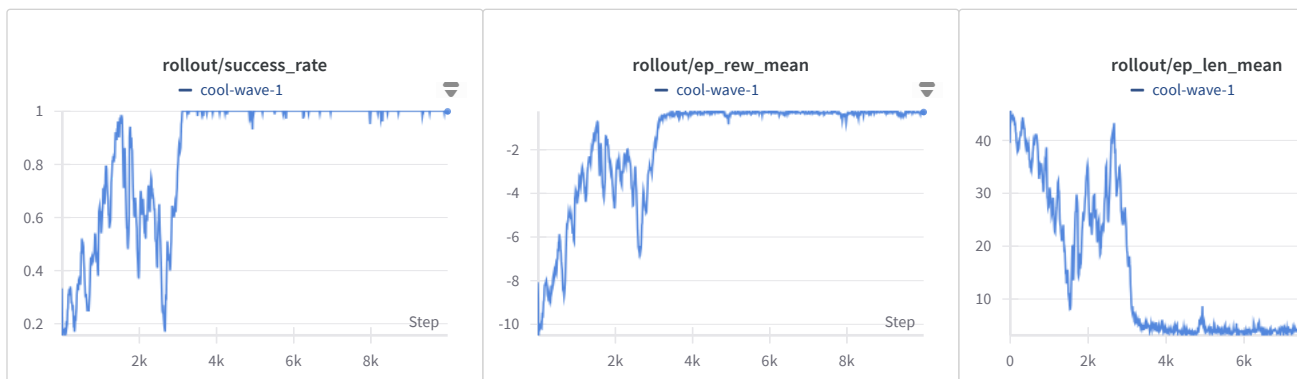


Import panel

Add panel



Ce graphique montre une progression régulière du nombre total d'étapes effectuées durant l'apprentissage. La montée linéaire indique que l'entraînement s'est déroulé de façon continue et stable jusqu'à environ 500 000 étapes.



Import panel

Add panel

**rollout/success_rate**

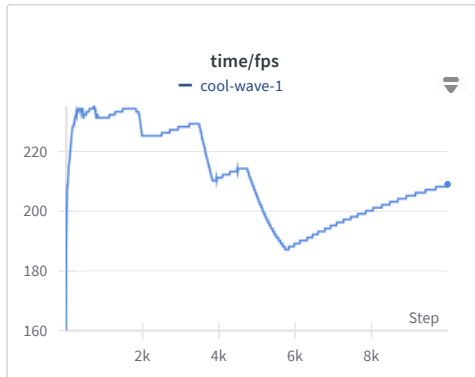
On observe une augmentation progressive du taux de réussite, qui commence bas et atteint une valeur proche de 1 après 4 000 étapes. Une fois ce seuil franchi, le succès reste élevé et stable, montrant que l'agent maîtrise bien sa tâche sur la suite de l'apprentissage.

rollout/ep_rew_mean

La récompense moyenne par épisode suit une évolution similaire : au début négative, elle s'améliore jusqu'à se stabiliser vers -2 après 4 000 étapes. Cela confirme que l'agent optimise son comportement au fil du temps, même si la récompense reste dans des valeurs négatives, ce qui peut être spécifique à la tâche.

rollout/ep_len_mean

La durée moyenne des épisodes est relativement élevée au début, avec beaucoup de variations, puis elle chute brusquement après 4 000 étapes pour se stabiliser à des valeurs très faibles. Cela indique que l'agent termine rapidement ses épisodes, ce qui peut refléter une efficacité accrue ou un comportement trop direct.

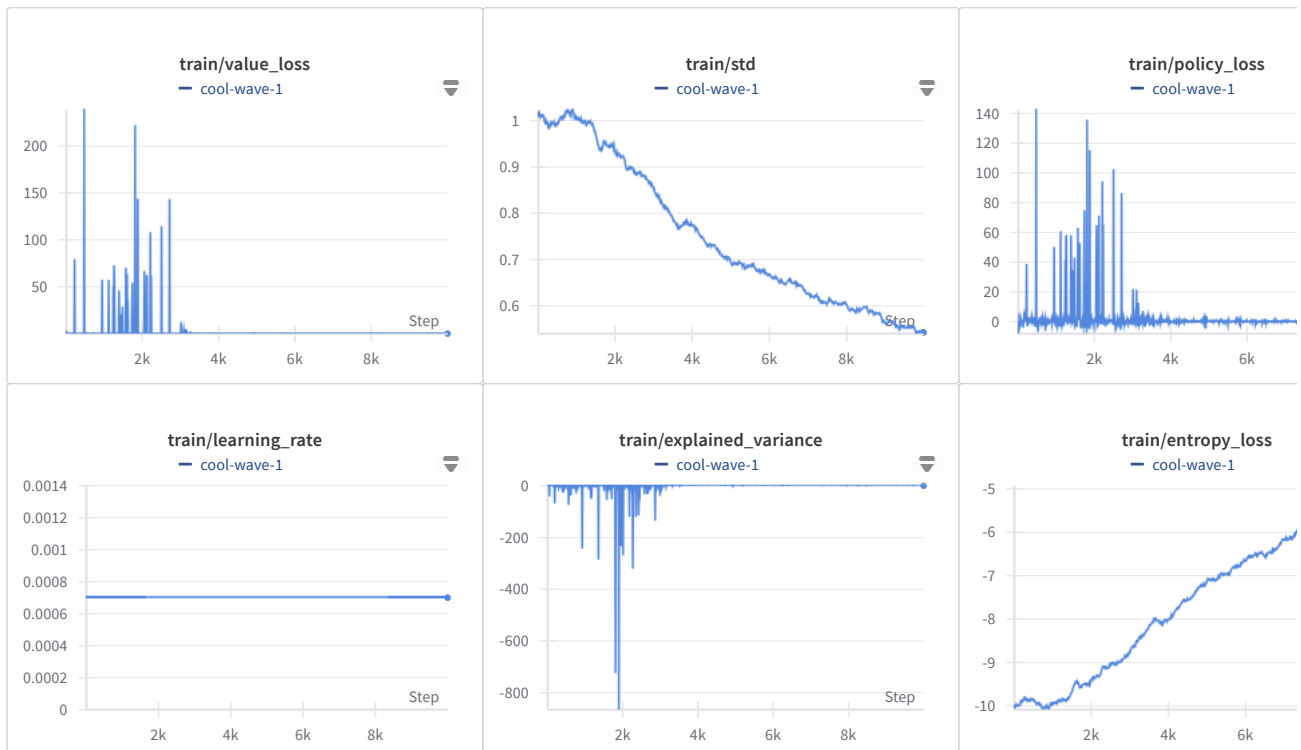


Import panel

Add panel



Le nombre d'images par seconde est stable au début autour de 220 FPS, puis diminue progressivement après 4 000 étapes, pour atteindre un minimum avant de remonter lentement. Cette variation peut s'expliquer par la charge de calcul ou les adaptations internes du modèle durant l'apprentissage.



Import panel

Add panel



train/value_loss

Au départ, on remarque de forts pics de la perte de valeur, ce qui est typique des premières phases d'entraînement. Ensuite, cette perte diminue fortement et devient quasi nulle après 3 000 étapes, ce qui reflète une bonne convergence de l'estimation de la valeur des états.

train/std

L'écart-type diminue progressivement tout au long de l'apprentissage, passant de 1 à environ 0,6. Cela montre que le modèle réduit la variabilité de ses prédictions, ce qui peut indiquer une plus grande confiance dans ses choix d'actions.

train/policy_loss

On constate une perte de politique très instable au début, avec de nombreux pics, mais cette instabilité disparaît progressivement. La courbe devient plate vers 3 000 étapes, traduisant une stabilisation de la politique d'action.

train/learning_rate

Le taux d'apprentissage reste constant autour de 0.0007 sur toute la durée de l'apprentissage. Cela assure une stabilité dans les mises à jour du modèle.

train/explained_variance

La variance expliquée est très fluctuante au départ avec des valeurs fortement négatives, ce qui témoigne d'une difficulté initiale à prédire la valeur des retours. Après 3 000 étapes, ces fluctuations diminuent et la variance reste basse mais stable.

train/entropy_loss

L'entropie augmente régulièrement tout au long de l'apprentissage. Cette progression indique que l'agent explore davantage avec le temps, ce qui favorise la découverte de nouvelles stratégies.